



**The American Psychologist**

Editor: **James H. Grier, Jr.**, University of Illinois

Contains original contributions of a theoretical nature  
Subscription: \$7.50 (Volume 17, 1962)

**Psychology of Man**

Editor: **John W. Berry**, University of Illinois

Contains studies of human behavior and thought  
Subscription: \$7.50 (Volume 17, 1962)

**Psychological Monographs**

Editor: **James H. Grier, Jr.**, University of Illinois

Contains monographs and shorter studies which  
Subscription: \$7.50 (Volume 17, 1962)

**Journal of Abnormal and Social Psychology**

Editor: **Charles W. Bailey**, University of Illinois

Contains experimental studies in the field of  
Subscription: \$7.50 (Volume 17, 1962)

**Journal of Experimental Psychology**

Editor: **Charles W. Bailey**, University of Illinois

Contains original contributions of an experimental nature  
Subscription: \$7.50 (Volume 17, 1962)

**Psychological Abstracts**

Editor: **Charles W. Bailey**, University of Illinois

Contains abstracts of the world's literature in psychology  
Subscription: \$7.50 (Volume 17, 1962)

**Journal of Applied Psychology**

Editor: **Donald G. Patterson**, University of Illinois

Contains material covering applications of psychology  
Subscription: \$7.50 (Volume 17, 1962)

**Applied Psychology**

Editor: **Donald G. Patterson**, University of Illinois

Contains studies of human behavior and thought  
Subscription: \$7.50 (Volume 17, 1962)

**Journal of Consulting Psychology**

Editor: **Donald G. Patterson**, University of Illinois

Contains studies of human behavior and thought  
Subscription: \$7.50 (Volume 17, 1962)

**The American Psychologist**

Editor: **James H. Grier, Jr.**, University of Illinois

Contains original contributions of a theoretical nature  
Subscription: \$7.50 (Volume 17, 1962)

**Journal of Comparative and Physiological Psychology**

Editor: **James H. Grier, Jr.**, University of Illinois

Contains original contributions in comparative  
Subscription: \$7.50 (Volume 17, 1962)

**American Psychological Association**

11 Dupont Circle, N.W., Washington, D.C. 20036

# Psychological Bulletin

---

## THE MASKING OF SPEECH\*

GEORGE A. MILLER

*Psycho-Acoustic Laboratory, Harvard University*

The ability to hear two sounds at once is one of the most useful properties of the human ear. This ability—noted by G. S. Ohm more than a century ago—allows us to respond selectively to certain components of the total acoustic pattern and to ignore others, to hear our friend's voice in spite of a background of noise, to follow the theme and still hear the obbligato. But, remarkable as this ability is, it is not infallible, and sometimes we are unable to hear one sound because another sound interferes. The selective mechanism is studied in its simplest form when two pure tones are introduced into the ear and the listener is asked to report the presence or absence of one or the other of the tones. In such experiments it is found that the ear is not a perfect analyzer, for some tones obscure the perception of others. This interference is called auditory masking.

Auditory masking is usually defined as "the shift of the threshold of audibility of the masked sound due to the presence of the masking sound" (23). Following this definition, the measurement of auditory masking is a straightforward experimental procedure. First, the just-audible sound-pressure is determined in the quiet. Then the interfering sound is introduced and the listener's threshold is again determined. The difference, in decibels, between the quiet and the masked thresholds is the measure of the amount of masking produced by that particular type and intensity of interference for that particular masked sound. The notion can be applied to complex noises like speech just as readily as it is applied to pure tones. The introduction of an interfering noise makes it necessary to raise the intensity of the speech for

\* Much of the research reviewed here was begun under an OSRD contract and is continuing under contract with the U. S. Navy, Office of Naval Research (Contract N5ori-76, Report PNR-23).

All of the research reviewed has been published previously but because of military classification some of the publications are available to only a limited psychological audience. In some cases it has been impossible even to refer to the original source.

it to be understood. The necessary increment is taken as the measure of the masking produced by the noise.

Knowledge of the ear's susceptibility to interference is obviously a matter of considerable practical value. Much of our present knowledge has grown out of the job of developing the telephone, although research on speech-communication problems in World War II has supplemented the earlier investigations. In the course of this research many different sounds have been studied to determine the interference they might produce, and a review of the results shows the masking of speech to depend on three characteristics of the masking sound: (1) its intensity relative to the intensity of the speech, (2) its acoustic spectrum, and (3) its temporal continuity. In the following pages we will review a variety of sounds which are, or might be, encountered, and in every case the disruption of vocal communication is determined by these three attributes. Human speech is most seriously masked by an uninterrupted noise which has its power concentrated in the lower third of a spectrum covering the frequency-range from 100 to 4000 or 5000 cycles.

#### THE SPECTRUM OF SPEECH

A noise which interferes with our perception of one sound may not interfere with our perception of another. Consequently, it is of some importance to examine the spectrum of human speech, since speech is the sound of direct concern for the present discussion.

Physically, speech consists of vibrations varying widely and rapidly in their intensity and frequency (2, 4, 19). In general, the vowels can be analyzed into discrete component frequencies, with fundamentals in the neighborhood of 100–200 cycles per second. Some consonants, however, contain energy distributed almost continuously through the high frequencies. Thus the energy is constantly shifting from one range of frequencies to another as the talker proceeds from one sound to the next.

One attempt to describe the energy in speech as a function of frequency borrows the concept of the spectrum from optics (2). A typical "long-interval speech spectrum" is shown in Fig. 1 (17). A crew of seven young men spoke an English sentence in a "conversational" voice, and a condenser microphone located 18 inches in front of the lips picked up the sound-waves and converted them into voltages which were analyzed by an audio-spectrometer. At a distance of 18 inches, the total pressure integrated over the entire range of frequencies was about 76 db above the standard reference pressure of 0.0002 dyne/cm<sup>2</sup>. The spectrum shown in Fig. 1 represents the root-mean-



square pressure in frequency-bands one cycle wide. For purposes of orientation the minimum audible field for pure tones is indicated at the bottom of the graph (18). This type of representation shows the long-time average distribution of speech-energy over the range of audible frequencies, but it is necessary to remember that the "instantaneous" spectrum is constantly shifting\* and is seldom, if ever, similar to the long-time average.

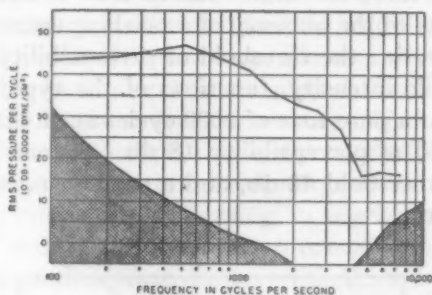


FIG. 1. THE LONG-INTERVAL SPECTRUM OF CONVERSATIONAL AMERICAN SPEECH FOR SEVEN MALE VOICES, EXPRESSED IN TERMS OF THE RMS PRESSURE IN FREQUENCY-BANDS ONE CYCLE WIDE.

The over-all level of the speech 18 inches in front of the talker's lips was 76 db re 0.0002 dyne/cm<sup>2</sup>.

In terms of a long-time average, therefore, speech is a low-frequency noise. Most of the power in the speech-wave is carried by frequencies below 1000 cycles. Thus, even in ignorance of the ear's susceptibility, the nature of speech itself suggests that communication is most seriously disrupted by sounds which have their energy concentrated in the frequencies below 1000 cycles.

#### ARTICULATION TESTING

When the over-all pressure of the speech-wave is less than about 5-10 db re 0.0002 dyne/cm<sup>2</sup>, the speech is inaudible. As the intensity of the speech is gradually increased the presence of spoken sounds can be detected even though none of the words is distinguishable. This intensity is conveniently referred to as the *threshold of detectability* for speech (3). Approximately 8 db above the threshold of detectability, the sounds begin to be perceived as words. If the spoken material is continuous discourse, it is possible, though difficult, to understand the gist of the

\* A recent attempt to include the temporal dimension is reported by R. K. Potter (15).

passage. This intensity is called the *threshold of perceptibility*. If the intensity is increased about 4 db more, the listener is able to obtain without perceptible effort the meaning of almost every sentence and phrase of the connected discourse. Above this level, the *threshold of intelligibility*, the intensity of the speech can be increased until it becomes painful at about 140 db (1). These thresholds are distinct and reliable, and different listeners will agree on their value.

Any one of the three thresholds can be used to determine the shift in the threshold due to the presence of a masking sound. With listeners instructed to determine the threshold of perceptibility in the presence of noises of variable intensity, functions of the type shown in Fig. 2 are obtained. Thus a pure tone of 1000 cycles at an intensity of 100 db shifts the threshold of perceptibility 18 db. A tone of 300 cycles at 100 db shifts the threshold 42 db, and a random noise of 100 db produces a 68-db shift.

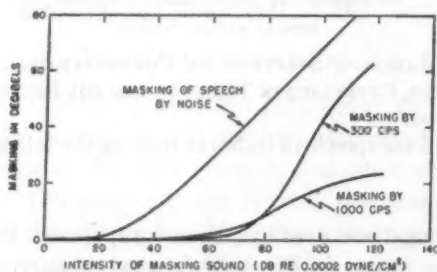


FIG. 2. THE SHIFT IN THE THRESHOLD OF PERCEPTIBILITY FOR SPEECH AS A FUNCTION OF THE INTENSITY OF DIFFERENT MASKING SOUNDS.

Random noise and pure tones having frequencies of 300 and 1000 cycles were used to mask the speech.

For a precise determination of the threshold, however, more elaborate methods are available. A talker reads out a list of discrete words, and listeners record what they hear. The percentage of the words heard correctly is taken as the articulation score. This method stands in the same relation to the threshold-methods as does the method of constant stimuli to the method of average error in classical psychophysics. By changing the intensity of either the speech or the masking sound a series of articulation scores ranging from 0% (no words heard) to 100% (all words heard correctly) is obtained. The 50% point on this function can be regarded as the threshold.

It is obvious that the results of these tests depend upon the type

of spoken material, the proficiency of talker and listeners, the characteristics of the equipment, etc. During the course of the development of articulation testing methods, from their first use by the Bell Telephone Laboratories (4) to their present wide application in communication problems, many variations have developed and considerable work has been devoted to their evaluation and standardization (3). The method adopted in the research reviewed here is recommended by its economy of time and personnel. Talkers are replaced by their recorded voices. Lists of difficult words are phonographically recorded in several different scramblings by each of five talkers. Between two and four trained listeners are used. As the records are played, the listeners follow the words with the aid of a check-list, uncovering each word *after* they hear it spoken. The listeners indicate by check-marks or manual counters whether or not they hear the word correctly. Since each listener must establish his own criterion for "hearing" a word, training is necessary before consistent results are obtained. A comparison with results from more formal articulation tests, however, indicates that with trained listeners the abbreviated procedure gives valid articulation scores.

It will be noted that the masking produced by various noises can be compared without obtaining the complete masking functions illustrated in Fig. 2. Single points on each function, if properly chosen, indicate the relative disruption of vocal communication which the different noises produce. Thus, if a quick check is desired, the threshold of perceptibility or intelligibility can be determined. If a more accurate value is needed, complete articulation functions can be run. One method is useful when a large number of variables need to be surveyed; the other gives more precise information when the significant variables have already been determined.

#### MASKING SPEECH BY TONES

Sounds which occur around us every day are customarily classified as either tones or noises. In this usage the word "tone" implies a harmonic relation between component frequencies. The word "noise" indicates that the component frequencies are dissonant, or randomly related. This is a classification of convenience, and a rigorous attempt to apply it to all the sounds which can be produced would quickly reveal borderline cases. Human speech, for example, is composed of sounds which have tonal characteristics (e.g. the vowels) and sounds with noisy characteristics (e.g. unvoiced plosives and fricatives), as well as many sounds which have both tonal and noisy components. In order to classify the variety of sounds which might interfere with

speech, at least three categories are required: (1) tones, (2) noises, and (3) voices. This rather arbitrary classification of the various sounds will serve to organize the following discussion.

*Pure tones.* As we would expect from the characteristics of speech, and as was indicated in Fig. 2, tones of low frequency produce more interference with speech than do tones of high frequency. This fact is supported by a very thorough study reported by Stevens, Miller and Truscott (21). These investigators found that for weak intensities the maximal masking is produced by sine waves in the vicinity of 500 cycles, whereas at high intensities the greatest masking occurs near 300 cycles. This effect is presumably due to the rapid spread of masking into the higher frequencies as the intensity of the masking tone is increased (22). This upward spread is aided by distortion in the ear which produces aural harmonics at high sound-levels (4, 20), and also by the characteristics of the response of the individual auditory-nerve fibers (9, 10, 13).

*Complex Tones.* Stevens, Miller and Truscott also determined the masking produced by square waves and by repetitive pulses of 10-microsecond duration. Square waves are less critical than sine waves as to frequency; fundamental frequencies between about 80 and 400 cycles mask with approximately equal effectiveness. No shift in the optimal frequency was observed as the intensity of the complex tones was increased. If the peak amplitude of the pulses is held constant, the greatest interference is produced by a pulse-repetition-frequency of about 200 pulses per second.

When the various wave-forms are equated for sound-pressure level, the sine wave is about 7 db less effective than the square wave, and the square wave about 7 db less effective than the pulses. This comparison is made for the low-frequency range in which the tones are most effective. At frequencies above about 1000 cycles all the tones are equally ineffective.

Before proceeding to discuss other types of interference we should pause to note two important generalizations supported by these data. First, low frequencies are more effective than high, and second, the greater the harmonic content of the tone the better its spectrum blankets the frequency-range of speech.

*Patterns of Tones.* Essentially the same considerations apply to the case of tones which are changing in frequency. Tests have been conducted with signals which changed abruptly or gradually in frequency, which covered wide or limited frequency-ranges at slow, intermediate or rapid rates, which varied in complexity or were accompanied by

steady tones, etc. In all the variations of tonal signals, the significant dimensions for masking were the frequency of the fundamental and the richness of the harmonic content.

Some examples should document these statements. One series of experiments was conducted with warbling tones—the tone rose slowly from the lowest to the highest frequency, then dropped suddenly back to the lowest frequency and rose again. The tone was produced by a relaxation oscillator and was rich in harmonics. Such a warble can be varied in center-frequency, range or rate. With the speech (heard binaurally

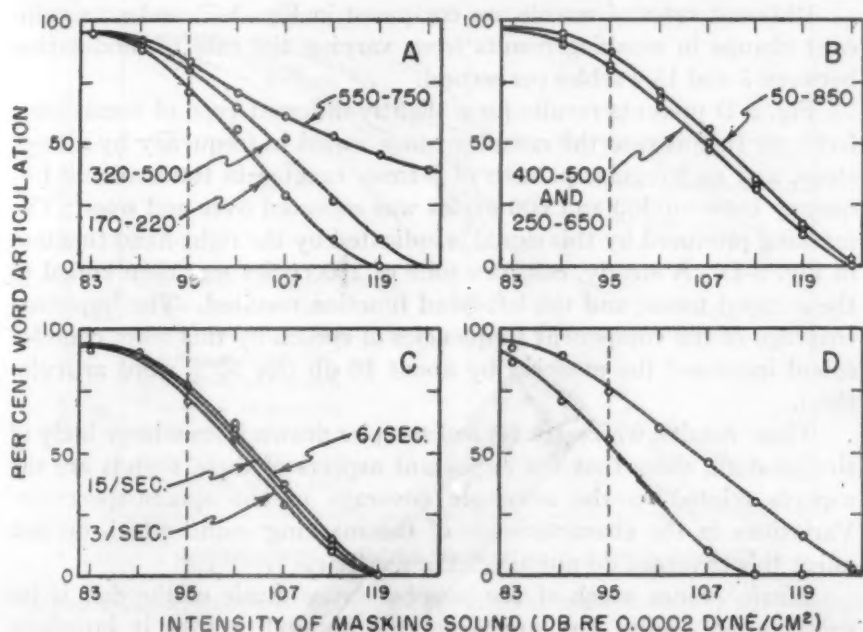


FIG. 3. THE PERCENT OF THE WORDS WHICH WERE CORRECTLY HEARD AS A FUNCTION OF THE INTENSITY OF VARIOUS TONAL MASKING SOUNDS.

Different center-frequencies of a warbling sound are compared in A; different ranges in B, different rates in C. In D the effect of adding a steady tone to a stepped pattern of tones is illustrated. The level of the speech was held constant at 95 db.

in dynamic earphones PDR-3) held at 95 db, the intensity of the warble was raised in 6-db steps until the functions shown in Fig. 3 had been obtained. Fig. 3-A compares warbles which had different center-frequencies. The high-pitched warble (550 to 750 cycles) resulted in the least interference and 50% of the words were still intelligible when



the warble was 20 db more intense than the speech. The low-pitched warble (170 to 220 cycles) interfered most, and 50% of the words were unintelligible when this sound was only 7 db more intense than the speech.

In Fig. 3-B three warbles are compared which had the same rate (1.5/sec.) and the same center-frequency (450 cycles), but which differed in the range of frequencies covered by the warbling tone. The widest range (50 to 850 cycles) was slightly less effective because it included some relatively high frequencies, but the differences are quite small.

Different rates of warble are compared in Fig. 3-C, and no significant change in masking results from varying the rate of modulation between 3 and 15 warbles per second.

Fig. 3-D presents results for a slightly different type of tonal interference. In this case the complex tones varied in frequency by abrupt steps, and an irregular pattern of 5 tones ranging in fundamental frequency between 300 and 600 cycles was repeated over and over. The masking produced by this signal is indicated by the right-hand function in Fig. 3-D. A steady, complex tone of 200 cycles was then added to the stepped tones, and the left-hand function resulted. The improved coverage of the component frequencies of speech by this more complex sound increased the masking by about 10 db (for 50% word articulation).

These results, which are typical samples drawn from a large body of similar data, show that the important aspects of these sounds are the aspects related to the adequate coverage of the speech-spectrum. Variations in the characteristics of the masking sound which do not affect this coverage do not affect the masking.

*Music.* Since much of the popular dance music of the day is (to some people) noisy and annoying, the possibility that it interferes seriously with speech was worth investigating. Listening tests quickly revealed, however, that most music is inoffensive for two reasons. Music is divided into phrases, and between phrases are pauses, and during pauses speech is intelligible. And in the second place, clarinet and trumpet solos usually fall above the range of frequencies which produces the most efficient masking of speech.

A very complex masking sound is obtained, however, if two or three phonographic recordings are played at the same time. The different orchestras fill in each other's pauses, and the coverage of speech-frequencies is more consistently adequate. With a signal of this type, an articulation function was obtained which was almost identical with the

function in Fig. 3-D for the complex stepped tones. Presumably, this function represents nearly the maximum masking which a tonal signal can produce. If the sound is made more complex it loses its tonal characteristics and begins to sound more and more like an irregular noise.

### MASKING BY NOISE

In many ways white noise, having a uniform, continuous spectrum, is the ideal sound to use in studying the masking of speech. This noise provides a continuous coverage of a wide range of frequencies, and the spectrum can be manipulated by filters to suit the manipulator's fancy.

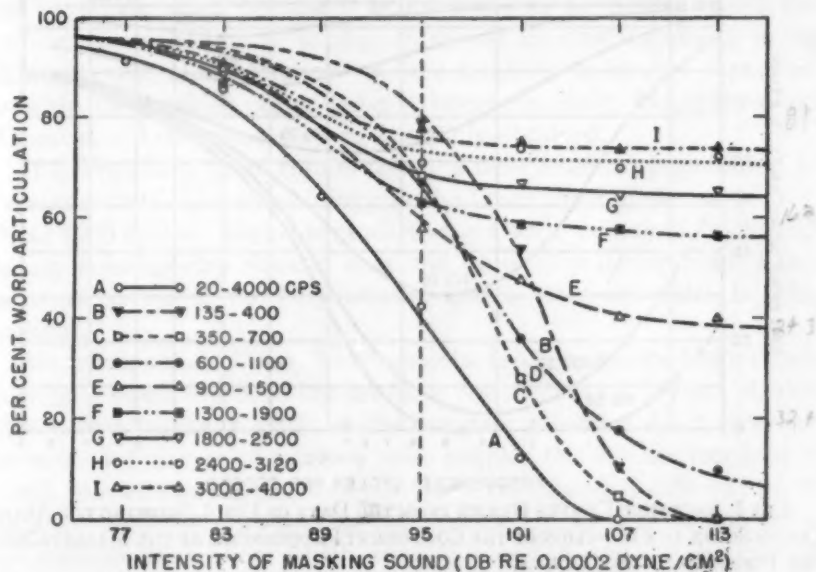


FIG. 4. THE ARTICULATION SCORE AS A FUNCTION OF THE INTENSITY OF THE MASKING NOISE.

The different frequency-bands of noise are parameters. The level of the speech was held constant at 95 db.

Thus we might use wide or narrow bands of noise, or noises "shaped" for maximal effect.

**Narrow Bands of Noise.** Let us review first some results obtained with narrow bands of noise. Highly discriminating filters (M-derived) were inserted into the noise-channel to provide adjacent pass-bands as follows: 135-400 cycles, 350-700 cycles, 600-1100 cycles, 900-1500 cycles, 1300-1900 cycles, 1800-2500 cycles, 2400-3120 cycles, and 3000-4000 cycles. These cut-off frequencies represent the frequency at which

the response of the system was 6 db below the maximum in the band. The narrow bands of noise were then mixed with the speech, which was held constant at 95 db, and articulation tests were run. The results are shown in Figs. 4 and 5. In Fig. 4 the articulation score is plotted as a function of the noise-level, with the different bands as parameters. In Fig. 5 the articulation score is plotted as a function of the center-frequency of the bands of noise (vertical divisions indicate the approximate cut-off frequencies of the bands), with the noise-level as the parameter.

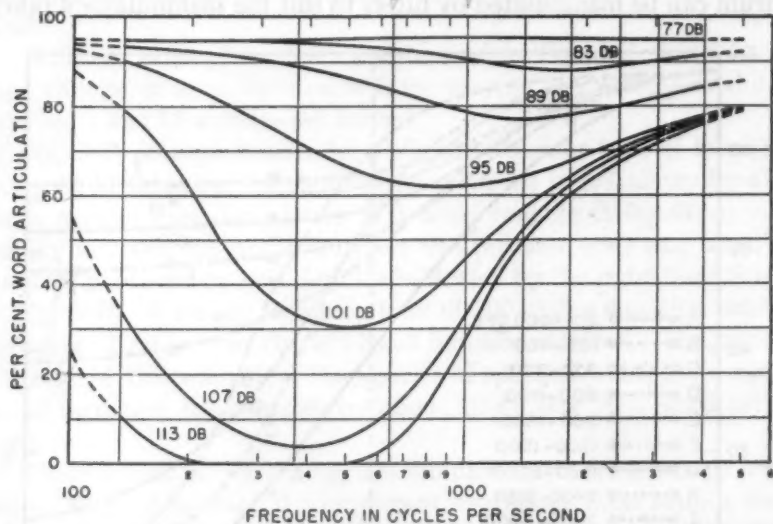


FIG. 5. SMOOTHED CURVES DRAWN FROM THE DATA OF FIG. 4, SHOWING THE ARTICULATION SCORE AS A FUNCTION OF THE COMPONENT FREQUENCIES OF THE MASKING NOISE FOR DIFFERENT NOISE-LEVELS.

Complete masking of the speech was obtained only with the two lowest bands. With bands of noise above 1000 cycles the speech could be heard even when the noise was more than 18 db above the level of the speech. Results with a wide band of noise (20 to 4000 cycles) are also shown in Fig. 4, and it will be noted that none of the narrow bands of noise produced more masking than the unfiltered noise. At the low noise-levels, the high-frequency bands were more effective than bands below 1000 cycles. At the high noise-levels, however, the low-frequency bands were more effective. This is seen most clearly in Fig. 5.

In order to interpret these results it is necessary to recall the results of studies on the masking of pure tones by pure tones. Sounds of low frequency will, if intense enough, eventually mask the entire range of

frequencies involved in speech, but the high-frequency sounds are not able to mask the low frequencies of speech. The fact that, at low noise-levels, the high-frequency bands of noise are more effective than the low is probably due to the distribution of energy in speech. The higher frequencies of speech (cf. Fig. 1) are much weaker and are therefore more easily masked. As the intensity of the high-frequency noise is raised, however, the high-frequency sounds of speech (mostly unvoiced consonants) are masked, but this masking does not spread to the low-frequency sounds. If, on the other hand, a band of noise composed of low frequencies is gradually increased in intensity, it will at first have little effect because it is not so strong as the low speech-frequencies, nor is it intense enough to spread its masking effects to the higher, weaker speech-sounds. As the intensity is further increased, however, the masking effects begin to cover the entire frequency-range of speech, and complete masking is rapidly obtained.

It follows from these results that the most masking is produced by a spectrum with the noise-energy in the lower third of the range from 100 to 4000 cycles. Experimentation with a wide variety of "tailored," sloping noise-spectra showed that the optimal masking noise has a spectrum similar to the long-interval speech-spectrum which is being masked.

*Modulation of the Noise.* Random noise is often described as a steady "hissing" sound which, at low levels, is not unpleasant to hear. Unlike intermittent bursts of static or the irregular clacking of a typewriter, the sustained nature of random noise enables the listener to adapt to it and, to a certain extent, to ignore it. Consequently, different methods were tried for changing or modulating the quality of the noise. One method used filters which were switched in and out irregularly. Another made use of the Sonovox, a device which can be substituted for the vocal cords as a sound-source for speech. With a Sonovox driven by a random noise voltage, the resulting "speech" resembles a hoarse breathy whisper.

The results with such noises need not be reviewed in detail. It is sufficient to say that the experimenters were again forced to the conclusion that the critical dimension of the masking noise is its spectrum. Modulations of the noise may affect the listener's comfort, for he will say he prefers to listen to the steady noise. But when he is driven to it, he will understand just as many words with a modulated masking noise as with a steady noise.

*The Masking Wave-Form.* Will two noises having identical spectra, but different wave-forms, produce the same masking effects?

In order to study this question a frequency-modulated (FM) audio oscillator was developed. This oscillator was modulated with noise, i.e., the frequency of oscillation was randomly varied. The resulting wave-form, shown schematically in the upper half of Fig. 6, resembles a sine wave of constant amplitude, but of varying frequency. Once the spectrum of the frequency-modulated noise had been measured, it was

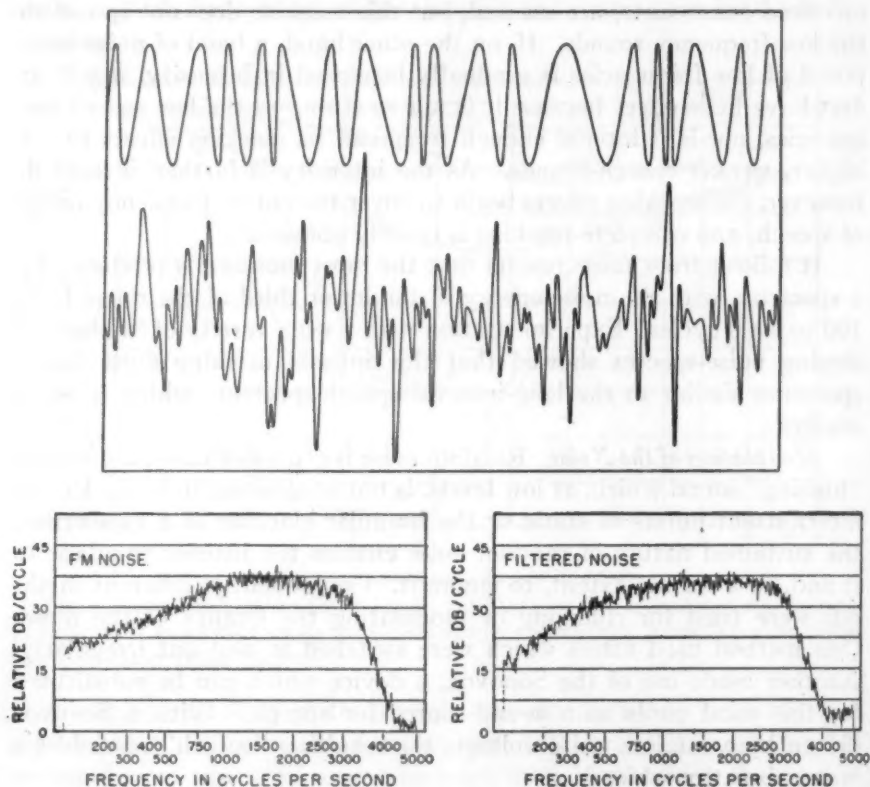


FIG. 6. FREQUENCY-MODULATED (FM) NOISE VS. RANDOM NOISE.

The wave-form at the top of the figure schematizes FM noise; the wave-form below, random noise. The matched spectra of FM and random noise are shown in the lower half of the figure.

then possible by the use of filters to "shape" the spectrum of a random noise of irregular amplitude in such a way as to make it correspond with the spectrum of the FM noise. When the two spectra shown in the lower half of Fig. 6 are compared, it is seen that the spectrum of the



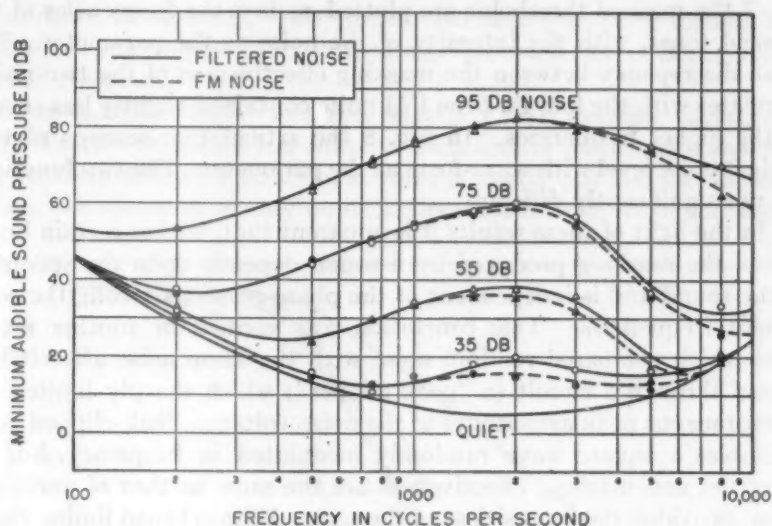


FIG. 7. THE SHIFT IN THE THRESHOLD OF AUDIBILITY PRODUCED BY TWO NOISES HAVING DIFFERENT WAVE-FORM BUT SIMILAR SPECTRA.  
(Cf. Fig. 6.)

FM noise is 2 or 3 db lower at frequencies above 2000 cycles. Otherwise the spectra are very similar.

The masking by these two wave-forms was then tested for both pure tones and speech, and the results are shown in Figs. 7 and 8. In

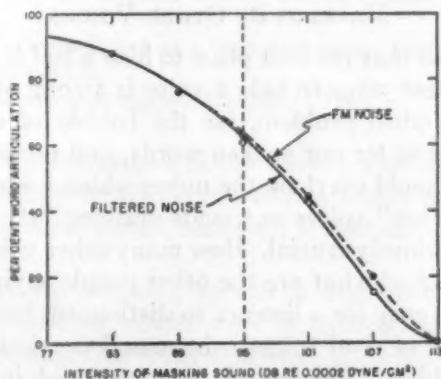


FIG. 8. THE ARTICULATION SCORE AS A FUNCTION OF THE INTENSITY OF TWO MASKING NOISES HAVING DIFFERENT WAVE-FORMS BUT SIMILAR SPECTRA.  
(Cf. Fig. 6.) The level of the speech was held constant at 95 db.

Fig. 7 the masked thresholds are plotted against the frequencies of the masked tones, with the intensity of the noise as the parameter. The small discrepancy between the masking effectiveness of the two noises correlates with the fact that the FM noise contained slightly less energy at the higher frequencies. In Fig. 8 the articulation score is plotted against noise level with wave-form as the parameter. The two functions are not significantly different.

In the light of these results it is apparent that, within certain broad limits, the masking produced by a sound depends upon the spectrum of the sound and is independent of the phase-relations among the component frequencies. This conclusion was verified in another set of tests which compared random noise with the same noise after it had passed through a circuit (a "peak clipper") which sharply limited the instantaneous peak amplitudes of the noise voltage. Peak-clipped noise resembles a square wave randomly modulated in frequency, but its spectrum and masking effectiveness are the same as that of unclipped noise, provided the intensities are the same. Within broad limits, therefore, the actual form of the masking wave as seen on a cathode-ray oscilloscope is of no consequence in determining masking. This agrees with the general hypothesis that the ear is relatively insensitive to phase-relations. The possible exceptions arise when the masking spectrum is produced by very intense but very intermittent bursts of noise. With such a sound the recovery-time of the ear becomes of appreciable importance, and we will have to consider specifically these temporal factors a few pages hence.

#### MASKING BY OTHER VOICES

It has been said that the best place to hide a leaf is in the forest, and presumably the best place to hide a voice is among other voices. Certainly it is a practical problem, for the babble of other voices is a frequent background for our spoken words, and no survey of common masking sounds should overlook the noises which we ourselves produce. In considering "other" voices as sounds masking "the" voice, two considerations are obviously crucial. How many other voices is the speaker competing with? And what are the other people saying?

It is relatively easy for a listener to distinguish between two voices, but as the number of rival voices is increased the desired speech is lost in the general jabber. This fact is demonstrated in the articulation functions of Fig. 9. The groups of interfering talkers were composed of an equal number of men and women (the single masking voice was a man's) reading and talking in a conversational tone. The "desired"

voices were, in all cases, male voices. For convenience in testing, the babble of voices was phonographically recorded by a high-fidelity recording system.

Note that the single voice is a relatively poor masking signal, and that even two voices are less effective than four or more. Although the long-interval spectrum of a single voice is nearly optimal for masking speech, the spectrum at any moment does not include all of the neces-

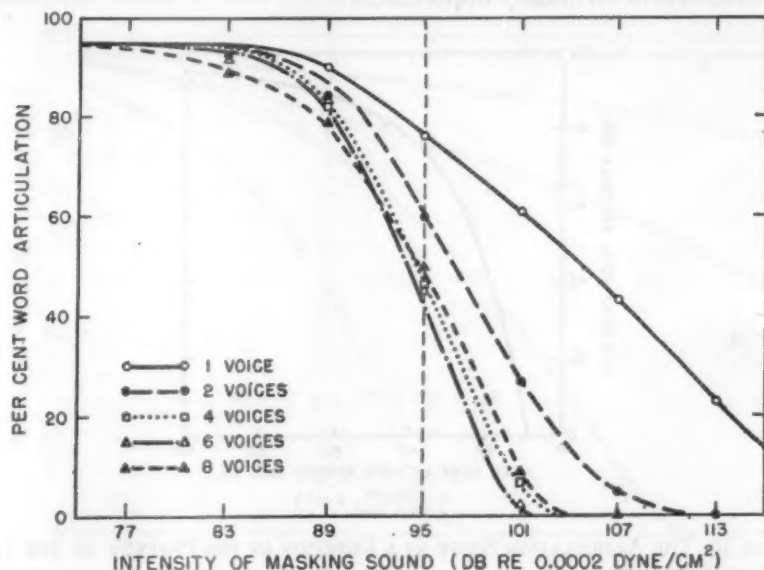


FIG. 9. THE ARTICULATION SCORE AS A FUNCTION OF THE INTENSITY OF DIFFERENT NUMBERS OF MASKING VOICES.

The level of the desired speech was held constant at 95 db.

sary frequencies. The variations in the level of a single voice are great, and there are relatively long intervals during which no masking sound is present. With several voices, however, a continuous masking signal is produced.

The content of the masking speech is a more difficult factor to evaluate. Conversational voices were compared with loud, excited voices liberally interspersed with laughter, cheering and improbable vocal effects. The two sounds could be likened to the chatter at a friendly dinner-party versus the din of a particularly riotous New Year's Eve celebration. There was little difference in masking, however. The shouting voices were a little more effective at low noise-levels, but

this was correlated with more high-frequency energy in the spectrum of the shouted babble.

Conversational babblings in different languages were also compared. A language was chosen which the listeners did not know, but the masking was neither greater nor less than was obtained with an English babble. Once again, it is necessary to conclude that the crucial factor is the masking spectrum. The particular way in which the spectrum is produced is of secondary importance.

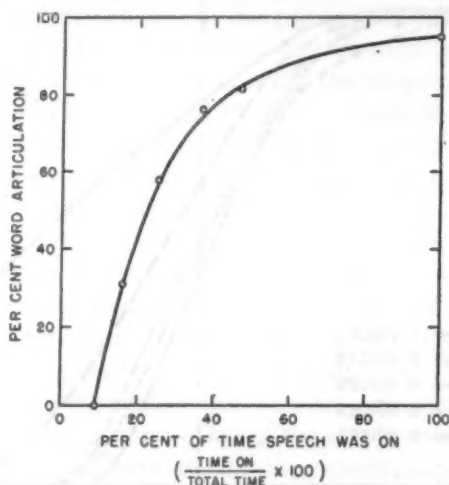


FIG. 10. THE ARTICULATION SCORE AS A FUNCTION OF THE PERCENT OF THE TIME THE SPEECH WAS ON.

The level of the speech was held constant at 95 db, and the speech-wave was interrupted at a rate of nine times per second.

#### TEMPORAL CONTINUITY OF MASKING SOUND

Not all of the sounds which force us to raise our voices are continuous. They come and go, change in quality and loudness. Consequently, some attention should be paid to this dimension of the masking sound. Because the ear is facile in patching together interrupted fragments of speech, an intermittent noise is not as serious a hazard as a continuous noise.

In order to demonstrate that large portions of the speech can be completely blanked out without seriously lowering intelligibility, an electronic switch was used to interrupt the speech nine times a second for variable portions of the total time. The function in Fig. 10 shows the relation between the intelligibility of words and the percentage of

the time the speech was present. With the speech on only half the time, 80% of the words were still correctly heard. It was not until 90% of the speech wave was missing that none of the words could be understood. (It should be noted, of course, that the function of Fig. 10 depends upon the rate of interruption and the type of test material, as well as upon the on-off ratio.) Qualitatively, the interrupted speech seems "hoarse" or "husky," as if the talker had some disorder of phonation. Nonetheless, the speech is surprisingly intelligible.

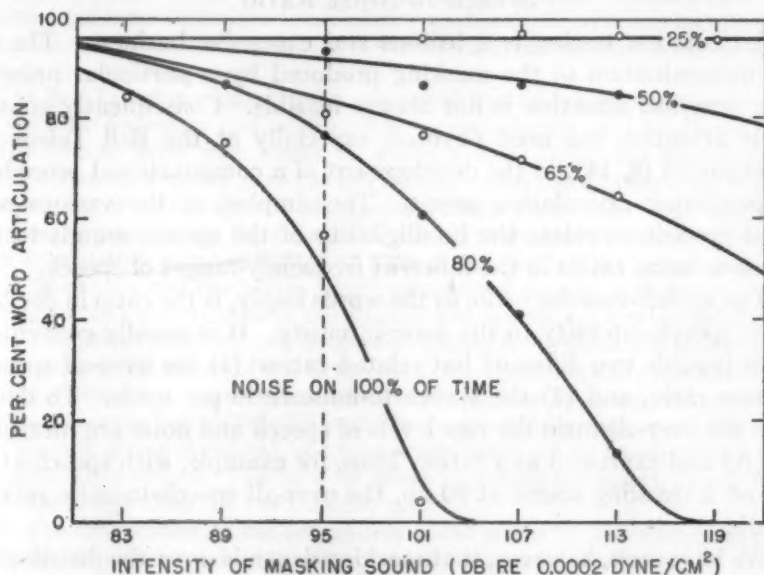


FIG. 11. THE ARTICULATION SCORE AS A FUNCTION OF THE INTENSITY OF INTERRUPTED MASKING NOISES.

The percent of the time the noise was on as parameter. The level of the speech was held constant at 95 db.

These results suggest that with an interrupted masking sound the ear should be able to integrate and interpret those portions of the speech available when the masking sound is not present. Consequently, the electronic switch was used to interrupt a masking noise instead of the speech. The results are shown in Fig. 11, where the articulation score is plotted against the intensity of the masking noise, with the percentage of the time the noise was on as the parameter. The noise-level in this case is expressed in terms of the intensity of the noise when it is on, and no average is taken between the level of the noise when it is on and the level when it is off.



When the noise is present less than 50% of the time, practically no masking is produced. Noise present 80% of the time is considerably less effective than uninterrupted noise. Apparently the recovery of the ear is rapid enough, and our ability to integrate fragments of speech is great enough, that any periodic interruption of a masking sound lowers its masking effectiveness. The careful documentation of this conclusion, however, remains for future investigators.

### SPEECH-TO-NOISE RATIO

Articulation testing is a tedious and expensive business. The direct determination of the masking produced by a particular noise in some practical situation is not always feasible. Consequently considerable attention has been devoted, especially at the Bell Telephone Laboratories (8, 14), to the development of a computational procedure for predicting articulation scores. The simplest of the various proposed procedures relate the intelligibility of the speech sounds to the speech-to-noise ratios in the different frequency-ranges of speech.

The speech-to-noise ratio, as the words imply, is the ratio in decibels of the speech-intensity to the noise-intensity. It is usually convenient to distinguish two different but related ratios: (1) the over-all speech-to-noise ratio, and (2) the speech-to-noise ratio per cycle. To determine the over-all ratio the rms levels of speech and noise are measured directly and expressed as a ratio. Thus, for example, with speech at 95 db and a masking sound at 90 db, the over-all speech-to-noise ratio is +5 db.

We have seen, however, that masking depends upon the distribution of the component energy over the frequency-range. Two noises with the same over-all level can produce quite different masking results depending upon the spectra of the noises. To estimate masking effectiveness, therefore, it is necessary to know the spectrum of the masking noise and the spectrum of the speech. With an audio-spectrometer the two spectra are determined. The ratio in decibels of the speech-level-per-cycle to the noise-level-per-cycle at any frequency is the speech-to-noise ratio for that frequency. Thus the speech-to-noise ratio per cycle varies with frequency, while the over-all ratio does not consider the spectral distribution of the two sounds.

Now in order to use speech-to-noise ratios in the computation of articulation scores, it is necessary to make three assumptions. These can be considered in order.

1. *Articulation scores depend upon the type of test-material used.* With any fixed speech-to-noise ratio, higher scores will be obtained with sentences than

with words, and the lowest scores will be obtained with meaningless syllables. It is desirable, therefore, to obtain some more fundamental measure independent of the particular test-material. Sentence, word, or syllable scores could then be related to this more fundamental index, and all three could be predicted directly once the fundamental measure had been computed. Such a fundamental measure has been proposed by the Bell Telephone Laboratories. They call it the *articulation index*, and its derivation leads us to the second assumption.

2. *The contribution to the articulation index made by any narrow band of speech frequencies is independent of the contribution made by other bands.* If the range of speech-frequencies could be divided into a number of narrow bands each contributing the same amount to intelligibility, and each contribution independent of all other contributions, the sum of the individual contributions could be taken as the articulation index.

The Bell Telephone Laboratories attempted to establish empirically the relation between frequency and intelligibility by articulation tests conducted with filtered speech. With male voices, for example, they found that frequencies of speech above 1660 cycles gave the same articulation score (68% with the test material they were using) as the frequencies below 1660 cycles. For that crew and those test-materials, therefore, an articulation score of 68% was equivalent to an articulation index of 0.50, and 1660 cycles divided the speech into two bands, one above and one below, which contributed equally, 0.50 apiece, to the total index. In a similar manner the range of frequencies was further subdivided until a complete function could be drawn relating the articulation index to frequency. Once this function was determined, the total range of speech frequencies was divided into a convenient number of bands, each contributing equally to the articulation index (19). When the intensity of the speech in all the bands is far enough above threshold, the total contribution of all bands gives an articulation index of 1.00. Just how far above threshold the speech in any band should be brings us to the third assumption.

3. *The contribution to the articulation index made by any narrow band of speech-frequencies depends upon the speech-to-noise-ratio in that band.* If the speech in a given band of frequencies falls far below the level of the noise in that band, the speech will make no contribution to the articulation index. It is necessary, therefore, to use a weighting value to express the fractional part of the maximum contribution the band can make. This weight depends directly upon the speech-to-noise ratio. When the speech is 30 db or more above the noise, the band makes its maximum contribution, and speech-to-noise values greater than 30 db are considered as equivalent to the maximum weighting.

The computational procedure evolving from this argument is straightforward. First we determine the spectra of the speech and of the noise. Then we divide the range of speech-frequencies into  $n$  bands. The weighting value for each band of frequencies is next computed from the speech-to-noise ratios in the bands. The maximum contribution of every band is multiplied by its weighting factor. And then the contributions of all the bands are totalled to give the articulation index. Finally, the index is converted into sentence, word, or syllable articulation scores as desired.

Obviously this description of the procedure is highly simplified, and many special cases arise. For example, when narrow bands of noise are used to mask speech, the procedure does not provide adequate estimates of the masking effectiveness of the low-frequency bands of noise. Suppose that we have divided the speech spectrum into 20 equivalent bands. According to the computational procedure outlined above, the narrow band of noise will mask out 4 or 5 of the 20 bands, but the remaining 15 or 16 bands will continue to contribute to the articulation index regardless of the noise-level in the masked bands. Articulation scores should, according to prediction, fall a few percentage points and then remain constant for higher and higher noise-levels. Actually, as Fig. 4 shows, this is the case for high-frequency bands of noise. For low-frequency bands, however, there is a shift of masking into the frequencies above the band, and this spread increases as the intensity is increased. A similar example could, of course, be drawn from the results obtained when pure tones are used to mask speech.

This particular difficulty can, perhaps, be encompassed in a computational procedure if the masked threshold is used instead of the noise-spectrum. But in order to restrict the process entirely to pencil and paper, it is necessary to be able to predict the masked threshold for tones from the spectrum of the noise. This prediction can be made satisfactorily for noises with continuous spectra (5, 6, 7), but the prediction for tones or narrow bands of noise cannot be made at present.

Another problem which was encountered in evaluating masking noises was the matter of temporal continuity. The procedure outlined does not consider interrupted noises, even though interruptions are crucial to masking effectiveness. If more were known of the build-up and recovery times of the ear, this dimension of masking could presumably be quantified in a satisfactory way. For the present, however, the temporal aspect of the masking process has been neglected.

In spite of these difficulties, a simple computational procedure has much to offer. As a first approximation its usefulness cannot be denied. The emphasis is placed upon the distribution of masking energy relative to the distribution of the energy in speech and a survey of masking noises confirms the belief that these spectral distributions are of central significance. Problems of wave-form, of patterning and modulation, of familiarity with the masking sound are ignored, and the results presented here indicate that these aspects of the masking sound are of negligible consequence.

The results of the research and of the discussion can therefore be briefly summarized: the greatest interference with vocal communication

is produced by an uninterrupted noise which provides a relatively constant speech-to-noise ratio over the entire range of frequencies involved in human speech. Unfortunately, most of the noises we compete with fill this general prescription.

#### THE QUESTION OF ANNOYANCE

It is all well and good to know about the masking effectiveness of unwanted sounds, but few of us spend much time in boiler-rooms, airplane-cockpits, or machine-shops, and when we do ride the subway we simply shut up until the car stops. Noises which we cannot shout out may be a hazard in some occupations, but most of us manage to get along fairly well in the average din of the average day. What really upsets our vocal communication is the sound that, although it does not shatter our eardrums, is a nuisance, that distracts or annoys us, or just "gets on our nerves." Every teacher knows the shudder that runs through a class when the chalk squeaks piercingly against the blackboard, and an otherwise amusing radio program can become an agonizing nuisance if we try to use the telephone.

When we take this problem into the laboratory, however, it seems to disappear right in front of our ears. The major difficulty rests in the fact that the listener's attitude is so important. If he is engaged in difficult mental work, it may be relatively easy to annoy him. But if he listens with a defiant attitude, any attempts to upset him with strange noises may prove more amusing than effective. And since most of the sounds we can use in the laboratory are out of context and relatively meaningless, the task of being successfully obnoxious is practically impossible. Annoyance depends primarily upon the particular listener and the particular situation in which he finds himself.

If, however, we are content to ignore some of the situational variables involved, it is possible to ask listeners to compare different sounds on the basis of their "annoyance value." Some simple listening situation is standardized and the listeners compare pairs, use a rating scale, or rank-order an array of sounds. One can then evaluate the variables contributing to annoyance value as defined by the situation, although the safety with which the results can be extended to other situations is open to question.

Listeners were presented with pairs of sounds and were instructed to indicate which of the two sounds was more annoying. In making this decision, the listeners were told to judge which sound of the pair would be more unendurable if they had to listen to it for a long period of time. These instructions, therefore, constitute the definition of annoyance.



Fairly consistent results were obtained with groups of 10 to 20 listeners, and the scale of annoyance constructed in this way agreed closely with results obtained with rating or rank-ordering procedures.

As an illustration, the results obtained with stepped patterns of tones will be considered. Eight different variables in the tonal pattern were studied for their effect upon annoyance-value.

1. The higher the pitch of the component tones, the greater the annoyance-value. The range of frequencies tested was from 200 to 1500 cycles.

2. A wide range of frequencies between the highest and lowest steps is more annoying than a restricted range. Listeners reported that the wide range of component frequencies tended to be perceived alternately, first as a complete pattern and then as two patterns, one of high and one of low pitch. This effect is very similar to figure-ground reversals in visual perception.

3. The addition of continuous tones to the stepped pattern of tones produces complex effects dependent upon the frequency-relation between the tones. Beats give the sound a rough pulsing irregularity which the listeners disliked.

4. Listeners asked to compare continuous sounds of different wave-shapes found the complex sounds, especially brief pulses, more annoying. In general, the sine wave was found to produce little annoyance.

5. Patterns of 3, 4, 6 and 12 tones were compared, but the number of different steps in the complete pattern had little effect on the judgments of annoyance.

6. If one of the steps of a pattern is slightly longer in duration than the others, a rhythmic quality is added which the listeners judged to be more annoying than tones of equal duration. Even more annoying, however, is the pattern in which all the tonal durations are randomly varying.

7. A slow rate of repetition for a pattern of tones is considered slightly more annoying than a rapid rate.

8. Up to a certain limit, the annoyance-value is increased if silent intervals are introduced between the successive steps.

These results typify listeners' responses to meaningless sounds. When meaningful sounds like speech or music were used, the listeners refused to apply the word "annoyance" in describing them. Annoyance did not seem to be a proper dimension of such sounds, but the listeners were agreeable to calling the sounds "distracting." Apparently, meaningful sounds have a higher "attention-value" than meaningless sounds.

These experimental results, supplemented by results obtained with other types of sound, indicate that annoyance is related to three aspects of the sound.

*Loudness.* The most important single factor in determining annoyance-judgments is the intensity of the sound. With sufficient intensity, any sound can be made annoying, and extremely loud sounds produce actual pain. Since this variable is so fundamental to annoyance, care was taken to equate the intensity of the signals when other aspects were being studied.

*Pitch.* In general, sounds having their energy concentrated among the higher



audible frequencies are more annoying than low-frequency noises (11, 16). In this respect, the frequency of the sound alters its annoyance-value in a manner opposed to the effect on masking. With a low-frequency noise we cannot hear speech, but with a high-frequency noise we are more apt to be annoyed.

*Modulation of Loudness and Pitch.* A third important factor is the modulation which the sound undergoes. Listeners report that they prefer to listen to continuous, unchanging sounds, and that a sound changing irregularly from moment to moment is more annoying than a sound which is changing regularly (12). Listeners feel that the distraction of a changing sound is less desirable than the boredom of a constant sound, and they retain this opinion even after many hours of articulation testing in the presence of different noises. Apparently the changes in loudness are more effective than changes in pitch, but the individual differences on this point are too conspicuous to permit a safe generalization.

These conclusions lead to the belief that what we are here considering as annoyance is a close relative to the problem of hedonic tone. Judgments of the pleasantness, indifference, and unpleasantness would probably have led to very similar conclusions. Thus, while the results may be interesting as an exercise in experimental esthetics, the character of the problem has somehow been altered by the experimental approach. The principal concern, it will be recalled, is with annoyance as a hazard to vocal communication. On this score the results are consistently negative, and at no point in the experimental results is there unequivocal evidence that the articulation scores obtained by trained listeners in the presence of an annoying sound were lower than the scores obtained in the presence of an indifferent sound which had the same acoustic spectrum. With the attitude adopted by listeners in the laboratory situation, annoyance is not a hazard to communication. And yet, sounds do differ in annoyance value, and annoyance or distraction does sometimes interrupt our verbal flow. Perhaps the most reasonable generalization, therefore, is that when a listener finds himself in a situation where he is vulnerable to auditory annoyance, he is most vulnerable to loud, high-pitched, unpredictable sounds. Just what situational and attitudinal factors contribute to his vulnerability, however, this research does not reveal.

#### SUMMARY

A wide variety of sounds have been investigated to determine the extent to which they interfere with vocal communication. The masking of speech has been determined by articulation testing methods, and estimates of annoyance have been obtained by the method of paired comparisons.

The sounds are classified as noises, tones, and voices. For all three

types of sound, the stimulus-dimensions determining both masking and annoyance are the intensity, the frequency or spectrum, and the temporal pattern of the sound. Masking depends primarily on the speech-to-noise ratio over the range of frequencies involved in speech. Sounds of low frequency mask this range more effectively than sounds of high frequency. Interruptions in the sound decrease the masking effectiveness.

Annoyance also increases as the intensity is raised, but low-frequency sounds are less annoying than high-frequency sounds, and intermittent, irregular sounds are more annoying than continuous sounds. There is no evidence, however, that annoyance interferes with vocal communications in the laboratory situation.

## BIBLIOGRAPHY

1. DAVIS, H., HUDGINS, C. V., MARQUIS, R. J., NICHOLS, R. H., PETERSON, G. E., ROSS, D. A., & STEVENS, S. S. The selection of hearing aids. *Laryngoscope*, 1946, 56, 85-115 (Report PNR-7 from the Psycho-Acoustic Laboratory, Harvard University).
2. DUNN, H. K., & WHITE, S. D. Statistical measurements on conversational speech. *J. Acoust. Soc. Amer.*, 1940, 11, 278-288.
3. EGAN, J. P. *Articulation testing methods II*. Psycho-Acoustic Laboratory, Harvard University, Nov. 1, 1944, OSRD Report No. 3802. (Available through the Office of Technical Services, U. S. Department of Commerce, Washington, D. C., PB 22848.)
4. FLETCHER, H. *Speech and hearing*. New York: D. van Nostrand, 1929.
5. FLETCHER, H., & MUNSON, W. A. Relation between loudness and masking. *J. Acoust. Soc. Amer.*, 1937, 9, 1-10.
6. FLETCHER, H. Loudness, masking and their relation to the hearing process and the problem of noise measurement. *J. Acoust. Soc. Amer.*, 1938, 9, 275-293.
7. FLETCHER, H. The mechanism of hearing as revealed through experiment on the masking effect of thermal noise. *Proc. nat. Acad. Sci., Wash.*, 1938, 24, 265-276.
8. FRENCH, N. R., & STEINBERG, J. C. Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. Amer.*, 1945, 17, 103.
9. GALAMBOS, R., & DAVIS, H. The response of single auditory-nerve fibers to acoustic stimulation. *J. Neurophysiol.*, 1943, 6, 39-58.
10. GALAMBOS, R., & DAVIS, H. Inhibition of activity in single auditory nerve fibers. *J. Neurophysiol.*, 1944, 7, 287-304.
11. LAIRD, D. A., & COYE, K. Psychological measurements of annoyance as related to pitch and loudness. *J. Acoust. Soc. Amer.*, 1930, 1, 158-163.
12. LAIRD, D. A. Influence of noise on production and fatigue, as related to pitch, sensation level and steadiness of the noise. *J. appl. Psychol.*, 1933, 17, 320-330.
13. LOWY, K. Some experimental evidence for peripheral auditory masking. *J. Acoust. Soc. Amer.*, 1945, 16, 197-202.
14. MUNSON, W. A. Relation between the theory of hearing and the interpretation of speech sounds. *J. Acoust. Soc. Amer.*, 1945, 17, 103.
15. POTTER, R. K. Visible patterns of

- sound. *Science*, 1945, 102, 463-470.
16. REESE, T. W., & KRYTER, K. D. *The relative annoyance produced by various bands of noise*. Psycho-Acoustic Laboratory, Harvard University, March 17, 1944, IC-65.
  17. RUDMOSE, H. W., et al. *Effects of high altitude on the human voice*. Electro-Acoustic Laboratory, Harvard University, Jan. 30, 1944, OSRD Report No. 3106. (Available through the Office of Technical Services, U. S. Department of Commerce, Washington, D. C., PB 19820.)
  18. SIVIAN, L. J., & WHITE, S. D. On minimum audible sound fields. *J. Acoust. Soc. Amer.*, 1933, 4, 288-321.
  19. STEINBERG, J. C., & FRENCH, N. R. The portrayal of visible speech. *J. Acoust. Soc. Amer.*, 1946, 18, 4-18.
  20. STEVENS, S. S., & DAVIS, H. *Hearing*. New York: John Wiley, 1938.
  21. STEVENS, S. S., MILLER, J., & TRUSCOTT, I. The masking of speech by sine waves, square waves, and regular and irregular pulses. *J. Acoust. Soc. Amer.*, in press. (Report PNR-14 from the Psycho-Acoustic Laboratory, Harvard University.)
  22. WEGEL, R. L., & LANE, C. E. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Phys. Rev.*, 1924, 23, 266-285.
  23. *American standard acoustical terminology*. American Standards Association. New York: Amer. Soc. Mech. Engineers, March 20, 1942.

## THE EFFECTIVE USE OF MANIPULATIVE TESTS IN INDUSTRY

W. F. LONG AND C. H. LAWSHE, JR.

*Division of Education and Applied Psychology, Purdue University*

"The practical possibilities of psychological tests are now generally conceded, both by the professional psychologist and the industrial layman. . . . The immediate future is likely to see a very extensive application of tests to industry. . . ." These statements were made by Henry C. Link (35) in 1919.

It is disturbing to realize that those words seem just as apropos today as they did twenty-seven years ago, and that there is little doubt that the use of tests in industry during the intervening period until now has not been as extensive as Link anticipated. What are the reasons for this slow development of industrial testing, and what are the possibilities of Link's "immediate future" with the extensive application of tests to industry being close at hand? A review of pertinent literature should uncover some answers to these questions.

Since most of the significant work in the field of industrial testing completed prior to about 1935 has been adequately summarized by others, notably Bingham (5), Viteles (50), and Garrett and Schneck (22), it is not necessary to include here reviews of work completed prior to that date. Obviously, even when that mass of data is eliminated from review, the volume of literature on all phases of industrial testing since 1935 is too vast for the present review; hence consideration is limited to the use of manipulative tests in business and industry, which area is probably representative of the entire field. In addition, the references reviewed are limited almost entirely to those which included objective validity data and, with the exception of a few reports of British investigations, only work done in this country is included. Certain related, general reviews and bibliographies of industrial testing by Tiffin (45), Sells (42), Zerga (52), and Benjamin (2), are recommended to those interested in a detailed coverage of the topic although none of them is included herein.

### EXTENT OF USE

To substantiate the suggestion in the second paragraph above that the use of psychological testing in industry is not nearly as extensive as it might be, certain reports can be noted.

In commenting about Link's statements mentioned previously, Taylor (43) noted in 1940 that the belief that the use of psychological

testing in industry is bound to increase, has been hopefully voiced from time to time throughout the intervening period of time since Link's article was published. He noted also that Hay (30) in the early part of 1940 listed eight companies which to his knowledge were making systematic and extensive use of tests in the selection of employees. Taylor then commented, "Granting that possibly two or three times as many other companies which did not happen to come to Mr. Hay's attention may have test programs, the number carrying on such activities certainly does not represent any sizable proportion of American industry." He suggested three reasons for the slow growth of testing programs: over-enthusiasm on the part of exponents of testing, over-skepticism by those unfamiliar with tests, and the expenditure of time and money required for the development of adequate industrial tests.

It should be noted that Taylor did not include in his commentary a part of Hay's remarks which said, "Many more companies have partial programs or are in process of developing a program, and it is probable that more serious attention is being given to psychological tests at the present time than ever before. Nevertheless it is remarkable, considering the success achieved by these companies, that every company in the country does not use psychological tests as a matter of course."

Some conflicting, but more convincing, testimony was given by Borow (10) when he reported that a National Industrial Conference Board survey made in 1936 found that slightly more than seven per cent of 2,412 firms used psychological tests. Although this is a small percentage of the total, the number is substantially larger than Hay or Taylor reported. Borow suggested in the same report that the number of companies using industrial tests in 1944 would dwarf the number listed in the National Industrial Conference Board survey, although no accurate figures were available to support his belief.

Judging on the basis of these few reports, there is no doubt that something is amiss in the use of psychological tests in industry, and that an explanation for this condition, which will at the same time suggest remedial measures, is needed.

#### SELECTION AND PLACEMENT

Most of the applications of testing in industry have been made in connection with the selection and placement of workers, probably because the most obvious use of tests is for those purposes, and because of the fact that management and workers can be more readily convinced of the value of tests for selection and placement than for other purposes.

In discussing occupational differences in manipulative abilities,



Teegarden (44) said, "A battery of manipulative tests makes possible rating of job applicants on specific traits of manipulative performance; speed, accuracy, delicacy of control, single or two-handed manipulation, ability to follow demonstration or instruction, to react to details, organize, maintain or increase speed, solve problems, etc."

He also reported that test performances of adults in 14 men's and 16 women's occupations show that no two occupations present identical combinations of score levels on the tests used, and mentioned the obvious conclusion that analysis of an applicant's test scores indicates the type of occupations for which he seems best fitted.

In their *Summary of Manual and Mechanical Ability Tests*, which includes objective descriptions of available mechanical aptitude tests and brief summaries of their application to various types of selection situations, Bennett and Cruikshank (3) wrote, "The assembly test or apparatus test which may be a miniature replica of the job situation, is probably a very adequate test for job selection, particularly for semi-skilled workers. No doubt much more adequate measures can be devised in this form of test."

It would be impractical to attempt to consider the use of each type of manipulative test separately, so a more logical division of test applications into occupational groups has been made.

*Machine Operators.* According to Hurt (32), in October of 1937 a garment factory in Marion, Virginia, started using four tests: Strength of grip, speed of arm movement determined by the length of time required to transfer 12 thimbles from one set of pegs to another and back again using one hand only, hand steadiness, which was scored by determining in how small an opening in a metal plate the subject could insert a stylus without making contact, and a general dexterity test. The latter test was considered to be the most valuable. It consisted of an inclined board with three equally spaced sockets in line at the front of the board with a counting device behind each socket. The subject punched the stylus successively in each hole as rapidly as possible for one minute. The average score was 35 and it was found that a score of 38 or better adequately predicted success of sewing machine operators. Although no quantitative validity data are presented, the author states that a few operators that were hired experimentally from the disqualified group demonstrated that they were not fitted for the job.

Using a miniature punch press in testing 25 experienced punch-press operators, Tiffin and Greenly (47) found that speed on the test is closely related to both speed and accuracy ratings made by foremen, the correlations being  $-.55$  and  $.63$  respectively. A high relationship

between accuracy on the test and a safety rating was also found to be present with a correlation of .60 reported.

Bennett and Fear (4) administered a series of tests to machine tool operators working on turret lathes, Bullard Automatics, precision grinders and milling machines at the plant of Martin and Schwartz, Inc. in Salisbury, Maryland. They used supervisor's ratings as a criterion and obtained a correlation of .40 with performance on the Wisconsin miniature test for the engine lathe, .41 with a hand-eye coordination test consisting of a dummy drill press, and .46 with a hand-tool dexterity test involving the use of a wrench and screwdriver. Another approach to demonstrate the usefulness of the tests was taken by the authors in reporting that 91% of the men rated excellent received excellent or good total test scores, 75% of the men rated good received scores of excellent or good, and none of the men rated below average or poor received a score of excellent.

A type of job-sample test used with sewing machine operators was reported by Blum (7). In taking the test, the applicant operated a sewing machine performing two tasks: (1) following an irregular line and (2) sewing between two parallel though abruptly irregular lines. A correlation of -.42 between time required to complete the test and production records was obtained. He proposed certain critical scores for further use. If a critical score of 260 were used, 20% of the poor group and none of the good group would be eliminated; with a score of 240, 28% of the poor group and 4% of the good would be eliminated; score of 220, 56% of poor and 20% of good would be eliminated; and with 200, 70% of poor and 28% of good would be eliminated.

An extension of the use of the Minnesota Rate of Manipulation Test is reported by Jurgensen (33). The Ziegler revision of the test was administered to a group of men hired to be converting machine operators in a paper mill. The performance level on this test depends upon speed of gross hand and arm movements. The men were divided into five groups by each of three supervisors on the basis of speed of work. The ratings thus obtained for each man were transformed into T-scores and the sum of the three scores for each man was used as a criterion. The maximum multiple correlation obtained was .60 for a group of 60 men.

Ross (38) found that when a critical score of 304 seconds on the O'Connor Finger Dexterity Test was used with a group of 41 machine-tool trainees, 90% of those rated "A," 100% of the "B's," 85% of the "C's," and none of the "D's" or "E's" would have been selected.

Andrews (1) reported the results of a year's experience with selection tests for engineering operatives. A battery of seven tests including in-

telligence tests and five apparatus tests was administered to 122 operators. Specifically, the apparatus tests included were a steadiness test, a finger dexterity test wherein small metal discs are placed as quickly as possible in recesses in a metal plate, a second dexterity test wherein pairs of screws under slight tension have to be unfastened and then fastened again, a bi-manual test involving simple coordinated movements of both hands, and a block sorting test made up of 100 blocks, thirty-six of which have part of a pattern design missing. It was found that if the highest scoring half of the applicant group had been hired, 87% would have been satisfactory, rather than the 60% which were satisfactory when selected without the use of tests.

*Assembly Workers.* Drake (16) used three tests, a pin board, a controlled turning test and a right-left turning test in experimental work for the Eagle Pencil Company. He found that those who scored above average on the test began training for a two-hand assembly job at 80% of normal worker efficiency and after two weeks training reached 97% efficiency. Those who scored below average on the tests began at 76% and ended at 91% of normal efficiency.

In a small group of watch factory workers Candee and Blum (11) found a correlation of only .26 between foremen's ratings and O'Connor Finger Dexterity Test scores. There was a critical ratio of 2.18 between mean scores on the test made by superior and mediocre workers.

Experience with employee selection tests for electrical fixture assemblers is reported by Tiffin and Greenly (46). In a group of 36 workers who performed operations of burning, twisting, and soldering ends of insulated wire a correlation of .63 was found between foremen's ratings of quality and scores on a hand precision test, and of -.16 between test scores and production records. They also reported a correlation of zero between ratings and finger dexterity test scores. For a group of 33 workers whose job consisted of placing plug and socket, soldering tips on ends of wires, fastening the wires and assembling the plug and socket the following relationships were reported. A correlation of .22 was found between production records and dexterity test scores (with experience held constant), and of .49 with hand precision test scores. Using supervisor's ratings of general efficiency as a criterion, a validity of .33 was obtained for the dexterity test (with experience held constant), and of .42 for the hand precision test. For a group of 44 operators whose work consisted of placing parts in a chassis and then connecting and soldering wires, the highest correlation found was only .27, that correlation being between dexterity scores and supervisor's ratings of efficiency (with experience held constant).

Blum (6) investigated the possibility of using the O'Connor Finger and Tweezer Dexterity Tests for selection of watch factory workers doing miniature assembly work. In a group of 137 workers, he found a correlation of .26 between finger dexterity time score and a salary ratio, .32 between tweezer dexterity time score and salary ratio, and .39 between a combined finger-tweezer dexterity score and salary ratio. He also reported a significant difference between time scores made by workers with less than seven days of experience and scores of workers with more than one year of experience, the latter group taking less time as might be expected.

*Inspectors.* Henshaw (31) used three performance tests with a group of 18 female paper inspectors. These girls inspected sheets of paper approximately 20×30 inches in size for weight and appearance by flipping the sheets over from one stack to another. All of the girls were 25 years of age or under and had at least six months of experience on the job. Using output figures as a criterion (reliability of .69 to .86) he obtained a correlation of .70 with a hand-arm dexterity test in which the subject continuously tapped two 3" keys located one foot apart one after the other. A validity of .61 was found for a tactile discrimination test in which the subject sorted a stack including three grades of paper. Using a choice-reaction test in which the subject reacted to one of three stimulus lights by pressing the correct one of three buttons to extinguish the light, a correlation of .52 was found with output records.

In studies of tests for the selection of inspector-packers, Ghiselli (23, 24) found the following correlations between a combination of ratings of supervisor and forelady: With a pegboard,  $-.50$ ; with the Minnesota Rate of Turning Test,  $-.40$ ; and the Minnesota Placing Test,  $-.24$ . The duties of the group of 26 girls consisted of (1) filling capsules, vials, and bottles with serums and antitoxins, (2) examining the filled containers for the presence of extraneous foreign matter, (3) labeling the containers, and (4) cartoning and packaging them.

*Other Industrial Workers.* The validity of certain mechanical ability tests for selecting cotton mill machine fixers is reported by Harrell (27). A tetrachoric correlation of .42 was obtained between scores made by 45 loom fixers on a 15 minute adaption of the Stenquist Mechanical Assembly Test and proficiency ratings made by supervisors. A correlation of .84 between scores made on the same test by 40 fixers in a carding department and ratings of mechanical ability made by an overseer was reported. The correlation between a composite of three ratings of mechanical ability and test scores of 10 spinning frame fixers was found to be .78. It is interesting to note the substantial differences between



the reported correlations when the criterion used was a rating of mechanical ability and when it was a rating of general job proficiency.

Drake and Oleen (20) reported the development and use of eight apparatus tests and one paper test in connection with studies for the Eagle Pencil Company. A comparison of the earnings of four groups differentiated on the basis of test scores is of interest. A group of 13 old employees with 32 months of experience who passed the test received wages in the amount of 116.1% of the average. A group of 10 employees with 24 months of experience who failed the tests reached a wage level of 93.6% of the average. A group of sixteen new employees who passed the tests and had only two months experience reached 97.4% of average level. A group of nine workers who passed the tests and received two weeks of training reached a level of 113.3% of average. They found that the cost of giving these tests was about two dollars per person which is less than half the difference in weekly wages between those passing the tests and those failing them.

A battery of tests including tests of motor coordination, intelligence, and emotional attitude was administered to 2,246 applicants for semi-skilled industrial jobs according to Cleeton (12). From this number, 849 applicants were selected for training, of which 546 successfully completed training. The only markedly significant difference between median test scores for the group which completed training and the group which was disqualified for apparent lack of ability was found for the motor coordination test. The median score for the first group was 75.5 and for the second, 61.3.

The use of certain undescribed machine and mechanical ability tests in the aviation industry is reported by Schultz (39). Of a group of 17 men rated "first class" by foremen, 88% received satisfactory test scores. In a group of 50 men rated as "semi-skilled," 88% also received satisfactory test scores. Only 42% of a group of 31 men rated as "bench-hands" received satisfactory scores. A year later 25 of 27 men selected by tests were doing good or average work.

Evans (21) found the median validity coefficient of a two-board, two-hand peg transfer test to be .34. The test was administered to 15 samples employed at work involving the use of two hands.

Cook (13) developed several tests on the basis of job analysis for the Western Electric Company. A coil winder test involving the task of winding wires in a prescribed manner around screws on a board differentiated a below average earning group from an above average group. In the above average group, 8% failed the test while in the below average group 72% failed. With relay adjustors, he used a monotony



test in which the worker tapped through a hole to hit a lower plate without hitting the sides of the hole in the top plate. In this group only 9% of those above average in efficiency were below average on the test, while 75% of those below average in efficiency were also below average on the test.

Moore (37) reported the use of manipulative tests in several situations. A Detroit company used the Detroit Assembly Test which consists of two fairly large boxes of 80 one-inch wooden cubes and requires the subject to use both hands to transfer the cubes from one box to the other at the same time packing the cubes neatly and in an orderly manner in the second box. When used in the selection of candy wrappers and packers, the company found that since the introduction of this test, 96% of the workers were satisfactory as compared to 52% before the test was used. He reported the use by the United States Employment Service of a wooden modification of the O'Connor Finger Dexterity Test in which the subject is required to move two pegs at a time, one in each hand, to a second board. A correlation of .45 was found between scores on this test and production records of 43 can packers. Moore also reported results of studies made by the United States Employment Service of several occupations using the Minnesota Spatial Relations Test. The following correlations were obtained between the test scores and production records or other factors: for can packers, from .20 to .24; for coding clerks, from .26 to .38; for calculating operators, from .32 to .59; for card punch machine operators, from .31 to .55; and for power sewing machine operators, from .40 to .50.

As reported in *Factory Management* (53), the Ford Motor Company used a hand dexterity test, a finger dexterity test, and a test wherein the subject is required to determine the size of rivets by touch rather than by sight. Foremen agreed that workers selected by the use of tests were superior to those selected by regular methods. In the same publication (54) it was reported that the Woodward Governor Company of Rockford, Illinois started using two tests of machine skill in 1938, and in 1943 the management stated that they were convinced that the tests are of inestimable value to the organization, not only for their value in placement but in upgrading. In one of the tests described, the subject controls a pencil by two cranks which turn in opposite directions and at different speeds and is required to follow a pattern on paper. In the other test described, the subject controls the path of a pencil by two cranks, following lines on a revolving drum. Although the two reports in this paragraph contain only subjective validity data, the tests and use made of them are especially interesting.

Crissey (14, 15) reported the use of tests in the AC Spark Plug Division of General Motors Corporation. A group of job-setters were given a number of paper and pencil tests, several well known apparatus tests, and two peg-board tests to check uni-lateral and bi-lateral dexterity. A validity of .65 for a selected battery is reported with supervisor's ratings used as a criterion. It was further reported that 69% of those who received test scores in the high third were also rated in the high third by supervisors, while none of those rated in the low third received scores in the high third of the group. A multiple correlation of .74 was found between a similar test battery and production records for aircraft spark plug gappers. The tests indicated that the most productive workers were above the average in bi-manual coordination and visual perception. In the same group, 88% of those who scored in the high third on the tests were also in the high third based on production records and all of those who had production records in the low third also received test scores in the low third. The tests also were demonstrated to be of value in the reduction of worker turnover rate when it was found that turnover for personal reasons of male employees for a one year period in a group pre-selected by tests was 5% while in a group not pre-selected the rate was 12%. He also presented one of the most convincing bits of evidence concerning the value of tests yet reported when he stated that as a consequence of hiring only those scoring in the high third on the battery of tests, the average production per operator was higher than the best previous individual record and there was considerable improvement in worker morale on the job.

As reported in *Occupations* (55), a battery of sixteen tests was administered to 51 aircraft riveter trainees. Those trainees whose total test scores were in the upper third of the total group correctly set 26% more rivets within a set period of time than those whose total test scores were in the lower third of the total group. The probability that such a difference would occur by chance is less than 1 in 1,000. Using the same timed work sample as a criterion, a multiple correlation of .60 was obtained for a battery of three tests which included a Worker-Analysis Pegboard Apparatus Part I, a Worker-Analysis Finger Dexterity Test Part II, and a figure-copying test.

*Department Store Workers.* Blum and Candee (8) investigated the possibility of using the Minnesota Placing Test, the Minnesota Turning Test and the O'Connor Finger Dexterity Test for selection of department store packers and wrappers. A multiple correlation of .38 was found between production records of a group of fifty-two seasonal employees and test scores while a similar correlation for permanent

workers was only .24. Efficiency on the job seemed to be the result of experience which was borne out by the fact that experienced workers scored significantly higher than seasonal workers, and that no difference in test scores was found between satisfactory and less satisfactory permanent employees. The authors suggested that these tests should not be used in selecting permanent employees.

In a later study, the same authors (9) found that manual dexterity as measured by the O'Connor Finger Dexterity and the Ziegler Placing Tests is not a selective factor for department store wrappers. With packers, a slight difference was found between the scores of experienced and inexperienced workers but this difference disappears with experience.

Ghiselli (25) found in checking the results reported by Blum and Candee that two of the same three tests plus a finger dexterity test would not be satisfactory as selection devices for package wrappers. In a group of 42 seasonal wrappers, the correlation of supervisor ratings with Minnesota Placing Test scores was  $-.10$ , with Minnesota Turning Test Scores  $-.02$ , and with the finger dexterity test scores  $.02$ . He suggested that if motor and dexterity tests are to be used as selective devices for this type of worker, they should be different in nature from the three used.

#### ACCIDENT PRONENESS

In beginning a study on the prediction and control of accidents, Drake (18) found that only moderate relationships had been reported between accident records and test scores when the tests had been handled in the conventional manner. He stated, "Far more significant relationships were found when the differences between scores on perceptual and motor tests were compared with an index that took account of both frequency and severity of accidents." The results obtained lead him to propose the hypothesis that accident proneness is a phenomenon associated with discrepancies in level between perception and motor reaction. It was observed that persons whose perceptual level is equal to or higher than their motor level are relatively safe, while those whose perceptual level is lower than their motor level are accident prone, with records of more frequent and more severe accidents than the former group. According to Drake, this implies that those who can see faster than they can react are relatively safe, while those who react faster than they can see are accident-prone. It is apparently possible however for the subjects studied to have a variety of uncorrected defects of vision and still get perceptual clues that were

quite adequate for effective and safe behavior. Drake noted that in order to make comparisons between perceptual and motor levels it was necessary to use tests that yield measures of the function intended and that are as free as possible from other complicating factors. In support of his contention that there is a marked tendency for difference scores to pick out workers with the high accident frequencies and severities, Drake offered as substantiating data, a comparison between accident indexes and difference scores. It is sufficient here to note that the half of a group of industrial workers that had the best difference scores had an accident index of 7.71 while the low half had an accident index of 25.04. He also reported that in one new group of employees selected on the basis of test and difference scores, accident reduction was 70% of the rate of workers selected by previous methods.

Williams (51) reported the use of a series of 15 simple tests in a study of accident prevention. The series included an interrupted pursuit test, a hand-eye coordination test involving dotting of small circles on a revolving disk moving at an increasing rate of speed, reaction to visual or auditory-tactual stimulus by pressing correct button, muscular hand-arm steadiness, and a hand dynamometer. Those scoring in the lowest quartile had an average accident history of .73 accidents per man per year while the average accident rate of the remaining 75% was only .37 per man per year.

### TRAINING

An interesting approach to the value of testing devices in training programs is reported by Tiffin and Lawshe (48). They found that hosiery mill employees with the poorest finger dexterity as measured by the Purdue Peg-Board cost a company \$59 each in minimum make-up before they made the rate, while employees having the best dexterity cost the company only \$36.40. They suggest that tests can answer three questions about training: (1). Who should be trained? (2). Where should training begin? (3). Has training been adequate?

Martin (36) reported the use of a battery of written and apparatus tests by the Woodward Governor Company. He stated that the company hadn't lost a member of the organization because of technical inability during the year since the tests were begun and that as a measure for weeding out the untrainable, the tests had worked out just about 85%.

Knowles (34) reported the experience of Northwest Airlines in using a battery of tests with general mechanics. The battery included one written test designed to measure ability to learn, a carefulness test in

which the subject is given a pile of metal pieces to assemble into units exactly like models, and a simple assembly test in which the subject assembles rows of bolts in a plate alternately from front and back while the plate is shielded from his view. In a group of 100 men hired for training, 25 were rated good, 50 fair, and 25 poor. If the 50 scoring highest on the test had been chosen, 38% would be in the good group, 58% fair, and only 4% poor. If the top 25 had been hired, 56% would have been in the good group, 44% fair, and none poor.

#### DEVELOPMENT OF TEST PROGRAMS

One of the striking impressions gained through reading published reports of industrial application of tests is the lack of ingenuity evidenced by many, if not most, investigators. The most successful test programs that have been reported included tests that were designed especially for specific situations. As could readily be discerned, however, the common procedure in too many instances was to get together a large group of tests that had worked in other places and administer them, hoping that some would prove to be of value. It also seemed that many times tests were administered and then the most expedient criterion was selected. Not disregarding certain advantages of standardized and ready-made tests, it seems painfully obvious that low validities in many situations were due to "shot-gun" test applications and careless selection of criteria. The resultant apparent limited value of many testing programs doesn't prove that tests are useless, but patently suggests that new tests should be developed and that any standardized tests used should be selected carefully on the basis of job analysis and possibly factor analyses, and that criteria should be chosen just as carefully.

*Improvement of Criteria.* In reviewing the literature the authors simply gained support for their own contention that one of the major problems in the effective use of tests in industry is the difficulty in obtaining adequate criteria. About eighty percent of the articles available for review were only generally descriptive and included no acceptable validity data. When given, validity data were often expressed in subjective terms, the usefulness thus being limited. Since all validation of tests, whether the methods used are simple or statistically complicated, depends on the availability and selection of adequate criteria, the first step in any development program must be the isolation of suitable criteria. To be most effective, a criterion must be reliable, relevant, and free from bias. In order to obtain criteria that meet these conditions, it is quite probable that new record systems of



production, employee ratings, absenteeism, accidents, tenure, training, and the like will have to be installed in many companies.

Drake (17, 19) suggested that all conventional criteria be discarded and proposed that test results be used as the criterion against which to measure the success of management, particularly the success of supervisors in releasing the measured abilities of workers in production. Apparently he found himself forced to make this proposal when he found that foremen's rankings, output figures, and efficiency ratings by time study techniques were proved to be unreliable as measures of operator ability on the job. This break with generally accepted procedure apparently ignores such considerations as motivation which would certainly be operating if the applicant felt he was setting his own performance level at the time of testing. Furthermore, a reliable criterion is only half of the solution since validity is at least equally as important as reliability.

*Job Analysis.* The necessity for the use of job analysis in developing a testing program for selection of mechanical workers was pointed out by Moore (37) when he wrote:

The level and type of mechanical ability that is needed depends upon job demands, and varies from one type of job to another. An assumption that is frequently made, and that has been as often disproved, is that there is such a unit as general mechanical ability. Mechanical ability can only be appraised when the job demands have been analyzed, interpreted, and converted into bodily elements. Therefore the selection of tests with which experimental work is to be carried on is determined by a careful analysis into human demands, a grouping of these demands in terms of functional similarities, and a preliminary selection of those tests which have been found valuable in appraising the same functions in similar situations.

Drake has written a number of articles in which he advocates the use of job analysis as the best possible basis for test development. In early studies reported with Oleen (20) the following job factors were considered useful in designing tests:

1. Length of cycle,
2. Nature of the elements of the cycle,
3. Sizes of materials and parts,
4. Serial order of elements of the cycle,
5. Three-dimensional position of parts manipulated,
6. Incidence of finger, wrist, arm, and body movements,
7. Posture of the operator,
8. Visual, tactual, and kinesthetic attentive factors,
9. Speed and rhythm of work.

When working with the Johnson and Johnson Company at New

Brunswick, New Jersey, to find suitable ways to measure observed differences in ability used on the job, Drake (17) found reason to comment, "We employed the methods then common to industrial psychology and still, I regret to say, too common in this field. We applied batteries of paper and pencil tests . . . and tried to find significant correlations between tests and various measures of success on the job such as foremen's ranks, production percentage efficiency, etc. This older procedure was cumbersome and costly and was characterized by trial and error to the extent that made it seem unscientific and illogical in the extreme." From the techniques developed in this study using time study analyses of jobs and groups of related jobs, six types of measurable human abilities were isolated. They were:

1. General finger, hand, wrist, arm dexterity,
2. Dual hand dexterity,
3. Bilateral hand dexterity,
4. Hand and foot coordination,
5. Machine tending ability,
6. Inspection ability.

Several performance tests were designed which had obvious similarities with the jobs studied and Drake reported that when these tests were administered again to the same operators after a lapse of time, sufficiently similar scores were obtained to indicate that the tests measured something relatively unchanging.

Cook (13) reported that he did not use detailed job analyses in selecting tests for his first work with the Western Electric Company, but followed the procedure of administering as many as 14 or 15 tests to the workers. He stated that the usual results were that two or three of the tests would prove to be of selective value. This procedure was considered to be wasteful of time and effort and consequently other procedures were investigated. Ultimately job analyses were developed which included nine factors as a basis for test selection. The factors included were intelligence, eye and hand coordination, finger dexterity, manual dexterity, small-tool dexterity, repetitiveness of work, accuracy, range of observation, and visual memory. Cook wrote that these factors were selected because it was felt that adequate tests were available for measuring an applicant's performance on any and all of them. Lest it be overlooked, it should be pointed out that the selection of job analysis factors on the basis of available tests makes use of only a portion of the potential value of the job analysis results to a test development program.

It seems probable that many psychologists working in industry who

have attempted to use job analyses in test development have not used them in the best possible manner. There has been too much of a tendency to designate abilities differentiated on the basis of superficial descriptions of the job or of worker abilities required to perform efficiently on the job. More thorough job analyses including the use of motion study to identify necessary bodily performance skills, and the identification of significant psychological traits and abilities such as visualization, numerical facility, and perceptual speed, should afford better bases for the selection and development of tests.

*Factor Analysis.* The use of factor analysis as a means of developing tests for industry is a forward step still to be used to the best possible advantage. Even this method cannot be utilized to its full potentiality, however, until more and better tests are included in the analyses.

Harrell (28, 29) reported a factor analysis of mechanical ability tests which included thirty variables, eleven of them apparatus tests including three pinboards, a pegboard, a peg sorting test, two tests of manipulating nuts and bolts, Crocket's block packing test, a test involving the placing of blocks along strips, a nut, bolt, and screwdriver manipulative test, and the wiggly blocks test. The tests were administered to a group of ninety-one cotton mill machine fixers. He found five factors which he named perception of detail (P), verbal relations (V), visualizing spatial relations (S), manual agility (A), and youth (Y). The tests which appeared in the agility factor were pinboard with both hands, two of the nut and bolt manipulation tests, block packing, pinboard with non-preferred hand, wiggly blocks, and a paper and pencil dotting test. Harrell noted that the P, V, and S factors have been previously identified by Thurstone and others. He suggested that since certain of the group paper and pencil tests measure each of the factors present in the manual tests, other paper tests can probably be devised to replace all of the manipulative tests. Considering the list of tests included by Harrell in the analysis, it seems highly improbable that much attention was paid to job analysis and that no attempt was made to develop tests especially to measure the abilities or aptitudes needed by machine fixers. He probably is correct when he suggests that paper and pencil tests could be devised to measure the abilities measured by manipulative tests that were included in the battery, but it is quite possible that better and more apropos manipulative tests could be developed that paper tests could not easily replace. It would also seem appropriate to question the necessity for the inclusion of tests for the measurement of mechanical abilities which are highly weighted with a

verbal factor, if equally valid manipulative tests could be developed that exclude a verbal factor.

Seashore (40) found that motor skill tests isolated into groups by factor analysis usually involved a similar pattern of movement regardless of musculature or sense-field required in subject performance which is a functional rather than an anatomical grouping of the skills. Later, Seashore in conjunction with Buxton and McCollum (41) suggested that the next step in identifying statistically located factors would be to use the technique of photographic motion study and simple introspection by the subjects rather than the all-too-frequent second hand analysis or vicarious introspection of the experimenters. They further suggested that any one physiological limit, such as speed or muscle contraction, is usually important only for a given work method and that changing to another work method may either partially or entirely overcome this limitation. This suggestion serves as one explanation of the success of applications of specially designed tests and the relative lack of success when shot-gun application of tests is made.

Guilford (26) makes some valuable suggestions in pointing out the results of certain factorial studies. He wrote:

Buxton's recent preliminary analysis tends to show that there are common motor factors. As Seashore points out, however, intercorrelations among motor tests are notoriously low and clusters are limited in scope. This fact forecasts a number of motor abilities of narrow range and tests with relatively low communalities. General factors of broad scope among motor abilities may include physical strength, agility, and steadiness of control.

The application of factor analysis to test development in industry is still rare, probably because the pressure to produce something immediately applicable does not often permit the use of a relatively long range program of factor analyses involving an accumulation of the necessary amount of intercorrelational data regarding tests and criteria. With sufficient data, the factorial content of a criterion can be determined and thus the criterion becomes a much more meaningful target for test development. One of the more obvious uses of the factorial approach to test development is to attempt to design pure tests to measure those unique factors that are demonstrated to be present in the criterion. As another result of factor analysis, the value of each test in predicting the known factors can be determined, thus permitting a close estimate of the validity of a test and the avoidance of the possibility of including more than one test in a battery which actually measure substantially the same thing. It is also possible that a particular test may be substantially loaded with a factor found to be related to success on a

job even though the test task has little apparent relation to job performance. The value of such a test probably would not be fully realized unless made evident as the result of factor analysis. It thus seems clear that factor analysis can assist considerably in making an industrial test research and development program effective.

### CONCLUSIONS

When using published material as a basis for judging the effectiveness of the use of psychological tests in business and industry, it must be remembered that many companies do not publish the results of their testing programs and it is quite probable that much successful and unsuccessful work remains unknown to all but a few. Nevertheless, a consideration of the published accounts of work done with manipulative tests would seem to indicate several conclusions in answer to the questions raised in the introduction of this review.

1. There is no doubt that testing programs can be most effective when the psychologists responsible for those programs are thoroughly familiar with industrial problems.

2. In many instances it may be necessary to redesign employee rating and record keeping systems to afford adequate criteria.

3. The effective combination and use of correct job and factor analyses in test selection and development will contribute greatly to the rapid development of extensive and effective psychological testing programs in industry.

4. The practice of reporting test results in other than technical terminology will probably contribute to the more ready acceptance of psychological testing programs by industrial management and workers.

5. Some of the most valuable applications of tests in industry may well be for purposes other than selection and placement of workers.

If the above suggestions are incorporated into psychological testing programs, there should be a much more rapid expansion of effective programs in business and industry than has been apparent to date.

### BIBLIOGRAPHY

1. ANDREWS, A. C. A year's experience of selection tests for engineering operatives. *Occup. Psychol.*, 1944, 18, 126-130.
2. BENJAMIN, H. C. Employment tests in industry and business: a selected, annotated bibliography. Industr. Relat. Sect., Princeton Univ., Bibliog. Series No. 67 (Rev.). Princeton, 1945.
3. BENNETT, GEORGE K., & CRUIKSHANK, RUTH M. *A summary of manual and mechanical ability tests (preliminary form)*. New York: The Psychological Corporation, 1942.
4. BENNETT, G. K., & FEAR, R. A. Mechanical comprehension and dexterity. *Person. J.*, 1943, 22, 12-17.
5. BINGHAM, W. V. *Aptitudes and aptitude testing*. New York: Harper, 1937.
6. BLUM, M. L. A contribution to man-



- ual aptitude measurement in industry: the value of certain dexterity measures for the selection of workers in a watch factory. *J. appl. Psychol.*, 1940, 24, 381-416.
7. BLUM, M. L. Selection of sewing machine operators. *J. appl. Psychol.*, 1943, 27, 35-40.
  8. BLUM, M. L., & CANDEE, BEATRICE. The selection of department store packers and wrappers with the aid of certain psychological tests, I. *J. appl. Psychol.*, 1941, 25, 76-85.
  9. BLUM, M. L., & CANDEE, BEATRICE. The selection of department store packers and wrappers with the aid of certain psychological tests, II. *J. appl. Psychol.*, 1941, 25, 291-299.
  10. BOROW, H. The growth and present status of occupational testing. *J. consult. Psychol.*, 1944, 8, 70-79.
  11. CANDEE, BEATRICE, & BLUM, M. L. Report of a study done in a watch factory. *J. appl. Psychol.*, 1937, 21, 572-582.
  12. CLEETON, G. U. Implications of industrial selection by tests. (Abstract.) *Psychol. Bull.*, 1938, 35, 678-679.
  13. COOK, D. W. Psychological tests for unskilled jobs. In *Psychological aids in the selection of workers*, Personnel Series No. 50. New York: American Management Association, 1941.
  14. CRISSEY, O. L. Tests predictive of success in occupation of job-setter. *Psychol. Bull.*, 1942, 39, 436.
  15. CRISSEY, O. L. Aptitude testing improves shop morale. *Amer. Mach.*, 1944, 88, 91-94.
  16. DRAKE, C. A. Spot the best by test; dual operations often effect time savings. *Fact. Mgmt.*, 1937, 95, 60-62.
  17. DRAKE, C. A. New developments in the selection of factory workers. In *Selection and development of foremen and workers*, Production Series No. 127. New York: American Management Association, 1940.
  18. DRAKE, C. A. The prediction and control of accidents. *Sci. Mon.*, N. Y., 1940, 51, 74-76.
  19. DRAKE, C. A. Job tests. *Person. J.*, 1941, 20, 184-189.
  20. DRAKE, C. A., & OLEEN, H. D. The technique of testing. *Fact. Mgmt.*, 1938, 96, 71-78.
  21. EVANS, S. D. A performance test for bi-manual jobs. (Abstract.) *Psychol. Bull.*, 1940, 37, 448-449.
  22. GARRETT, H. E., & SCHNECK, M. R. *Psychological tests, methods, and results*. New York: Harper, 1933.
  23. GHISELLI, E. E. Tests for the selection of inspector-packers. (Abstract.) *Psychol. Bull.*, 1941, 38, 735.
  24. GHISELLI, E. E. Tests for the selection of inspector-packers. *J. appl. Psychol.*, 1942, 26, 468-476.
  25. GHISELLI, E. E. The use of the Minnesota Rate of Manipulation and the O'Connor Finger Dexterity Tests in the selection of package wrappers. *J. appl. Psychol.*, 1943, 27, 33-34.
  26. GUILFORD, J. P. Human abilities. *Psychol. Rev.*, 1940, 47, 367-394.
  27. HARRELL, W. The validity of certain mechanical ability tests for selecting cotton mill machine fixers. *J. soc. Psychol.*, 1937, 1, 279-282.
  28. HARRELL, W. A factor analysis of mechanical ability tests. (Abstract.) *Psychol. Bull.*, 1939, 36, 524.
  29. HARRELL, W. A factor analysis of mechanical ability tests. *Psychometrika*, 1940, 5, 17-33.
  30. HAY, E. N. The use of psychological tests in selection and promotion. *Personnel*, 1940, 16, 114-123.
  31. HENSHAW, L. S. Selection tests for paper sorters. *Hum. Factor, Lond.*, 1937, 11, 146-153.
  32. HURT, J. Evaluating applicants by dexterity testing. *Employ. Serv. News*, 1939, 6, No. 6, 7-8.
  33. JURGENSEN, CLIFFORD E. Extension

- of the Minnesota Rate of Manipulation test. *J. appl. Psychol.*, 1943, 27, 164-169.
34. KNOWLES, W. We've stopped hunch hiring. *Nation's Bus.*, 1945, 33, 23-24.
  35. LINK, H. C. *Employment psychology*. New York: Macmillan, 1919.
  36. MARTIN, I. C. How and why Woodward Governor Company uses aptitude tests for employees. *Sales Mgmt.*, 1941, 48, 18-20.
  37. MOORE, H. *Experience with employment tests*. New York: National Industrial Conference Board, 1941.
  38. ROSS, L. W. Results of testing machine-tool trainees. *Person. J.*, 1943, 21, 363-367.
  39. SCHULTZ, R. S. Personnel selection in the aviation industry. *Person. J.*, 1940, 19, 99-105.
  40. SEASHORE, R. H. An experimental and theoretical analysis of fine motor skills. *Amer. J. Psychol.*, 1940, 53, 86-98.
  41. SEASHORE, R. H., BUXTON, C. E., & MCCOLLOM, I. N. Multiple factorial analyses of fine motor skills. *Amer. J. Psychol.*, 1940, 53, 251-259.
  42. SELLS, S. B. Measurement and prediction of special abilities. *Rev. Educ. Res.*, 1944, 14, 38-54.
  43. TAYLOR, H. C. A test program in an industrial company. In *Values of psychology in industrial management*, Personnel Series No. 43. New York: American Management Association, 1940.
  44. TEEGARDEN, L. Occupational differences in manipulative abilities. *Psychol. Bull.*, 1936, 33, 736.
  45. TIFFIN, J. *Industrial psychology*. New York: Prentice-Hall, 1943.
  46. TIFFIN, J., & GREENLY, R. J. Employee selection tests for electrical fixture assemblers. *J. appl. Psychol.*, 1939, 23, 240-263.
  47. TIFFIN, J., & GREENLY, R. J. Experiments in the operation of a punch press. *J. appl. Psychol.*, 1939, 23, 450-460.
  48. TIFFIN, J., & LAWSHE, C. H., JR. How tests can strengthen the training program. *Fact. Mgmt.*, 1944, 102, 119-121.
  49. VITELES, M. S. Psychology in industry. *Psychol. Bull.*, 1930, 27, 567-635.
  50. VITELES, M. S. *Industrial psychology*. New York: W. W. Norton, 1934.
  51. WILLIAMS, A. W. Accident prone workers can be weeded out by simple tests. *Textile World*, 1943, 93, 68-69.
  52. ZERGA, J. E. The development and use of apparatus tests in industry. *J. appl. Psychol.*, 1944, 28, 199-202.
  53. *How three companies use tests to select and place people*. [Anon.] *Fact. Mgmt.*, 1943, 101, 113-120.
  54. *Use the right selection and testing methods*. [Anon.] *Fact. Mgmt.*, 1943, 101, 96-105.
  55. *Ten years of occupational research*. Division of Occupational Analysis and Manning Tables, Bureau of Manpower Utilization, War Manpower Commission. *Occupations*, 1944, 22, 387-446.

## THE TREATMENT OF QUALITATIVE DATA BY "SCALE ANALYSIS"

LEON FESTINGER

*Research Center for Group Dynamics, Massachusetts Institute of Technology*

The basic problem of measurement in connection with measuring instruments such as questionnaires and interviews is to arrange a number of individuals in rank order from high to low on the basis of their responses to a series of questions. The solution of this measurement problem, that is, being able to say that one individual is higher, lower, or equal to another individual, along some specified dimension, has been attempted in a variety of ways. Dependence upon expert ratings, examination of intercorrelations among items and correlations between individual items and total scale scores are some of the methods which have been used to facilitate scale construction.

During the war Dr. Guttman, together with others in the Research Branch of the Information and Education Division of the War Department, developed a new technique for analyzing qualitative data which they have called "scale analysis" (2, 3, 4, 12). This technique attempts to solve some of the problems of scale construction which have up to now not been adequately handled.

There have as yet been relatively few publications concerning "scale analysis." The technique has, however, received quite a bit of favorable attention from social scientists. McNemar (7), for example, in a recent article make the following statements:

The next step would be to construct a uni-dimensional scale for measuring each of several components. Use of the Guttman technique would greatly facilitate the meeting of this requirement. . . . A Guttman-type scale for each component would also avoid the absurdity of having several questions, which supposedly tap a given component turn out . . . to have little or nothing in common.

. . . we have stressed the basic need for reliability, validity, and uni-dimensionality for the instruments or devices used to classify or measure individuals with respect to their opinions or attitudes. . . . Unitary scales can be developed by the Guttman scaling technique.

It is the purpose of this paper to review the published literature and some unpublished material with reference to the theory of "scale analysis," techniques of scale construction using "scale analysis," and the evaluation and interpretation of the scales achieved by this method.

## THE THEORY OF "SCALE ANALYSIS"

*The meaning of uni-dimensionality.* When an investigator attempts to measure a given variable (e.g. attitudes toward Russia, intelligence, weight) he assumes, at least implicitly, that he is dealing with something which has a unitary character. That is, that he can obtain measurements which relate unequivocally to this variable.

If the concept in question, or if the measuring instrument used, is not of this type, then there will unavoidably be some equivocation in the use of the measurement scale derived. To take a very simple illustration, suppose we were measuring individuals with an instrument (e.g., a series of questions) which simultaneously measured prejudice toward Negroes and mathematical ability. Two individuals who received the same score on this measuring instrument need not at all be alike. One could be more prejudiced and the other more able mathematically. Similarly, if one individual gets a higher score than another individual, we would be uncertain to what to attribute the difference. The difference might be one of degree or it might be a difference in kind.

In any measuring instrument which is not uni-dimensional, that is, which does not measure one and only one thing, the ordering of individuals by virtue of their scores does not have the simple quantitative properties which we desire in a scale. On an intelligence test, for example two individuals may achieve the identical score in very different ways. One individual might have good mathematical ability and another individual good verbal ability. These two quite different individuals nevertheless may come out with the same total score. In what sense then can it be said that the two are equal in intelligence or in what sense can it be said that a person who gets a higher score has more intelligence than a person who gets a lower score? The differences which we obtain between individuals may be differences in type of intelligence or type of ability, in addition to differences in the degree to which they possess these abilities.

The conclusion should not, of course, be drawn that the scores on intelligence tests, or the scores on any other measuring instrument, which does not possess uni-dimensionality, have no meaning. On the contrary, they can still be useful and have a good deal of meaning (1). It can be said, however, that if such a test could be split into several components, each of which did possess uni-dimensionality, then the value of this instrument would be greater than the value of the previous test.

Where it is feasible, the attempt should be made to have our measuring instruments measure only one dimension or one variable at a time.

Even if this is not possible, it is obviously desirable that the investigator have some indication of the extent to which his measuring instrument departs from the ideal of uni-dimensionality.

*The determination of uni-dimensionality.* The problem still remains of determining whether or not a given measuring instrument possesses uni-dimensionality. Let us examine the characteristics which such a measuring instrument would have.

It is implied by the preceding discussion that a uni-dimensional measuring instrument would be such that a given score on that instrument could be obtained only from one pattern of responses. In addition, for these scores to represent an ordering of individuals with respect to some variable there would have to be a certain type of consistency among the responses to the various items making up the measuring instrument (4, 11, 12).

Let us examine what the data would look like if secured with a measuring instrument known to be uni-dimensional. Suppose that in order to measure the height of individuals a measuring instrument were chosen which consisted of ten sticks, each one of different length. The operation of measurement is to stand each stick up alongside each individual and to record whether he is taller or shorter than the stick. We record a plus if he is taller and a minus if he is shorter. Our data for each individual then consist of ten marks, each of which may be either plus or minus.

We might now determine how many individuals received a plus on each of the measuring sticks. We could then arrange our sticks in order from the one where the greatest number of people received a plus down to that stick where the smallest number of people received a plus.

We would now find that any individual who received a minus on stick "1" would have received minuses on all of the other sticks; an individual who received a plus on stick "1" and a minus on stick "2" would have received minuses on all of the subsequent sticks and so on. The individual who had received a plus on stick "10" would have received pluses on all of the other sticks. Each individual would thus have fallen into any one of 11 "patterns of response." These patterns could be arranged in order from "tallest" to "shortest" and we would thus have a scale for measuring height. It may be noted that since this instrument was uni-dimensional, it would have been impossible for an individual, for example, to have received a minus on stick "1," a plus on stick "2," and a minus on stick "3." In fact, there are 1013 possible "patterns of response" which would *not* have occurred.

If our ten measuring sticks were ten questions which attempted to



determine, for example, attitudes toward labor, the same type of consistency would evidence itself if uni-dimensionality were present.

If then, after having collected data, we find them to present this type of consistency, we may conclude that our measuring instrument possesses uni-dimensionality. It should be emphasized that whether a given set of questions (a given measuring instrument) does or does not possess uni-dimensionality is entirely a matter for empirical determination.

#### THE TECHNIQUE OF SCALE ANALYSIS

*The Scalogram Board.* The scalogram board technique for executing a "scale analysis" was devised by Guttman (4). It rests upon the fol-

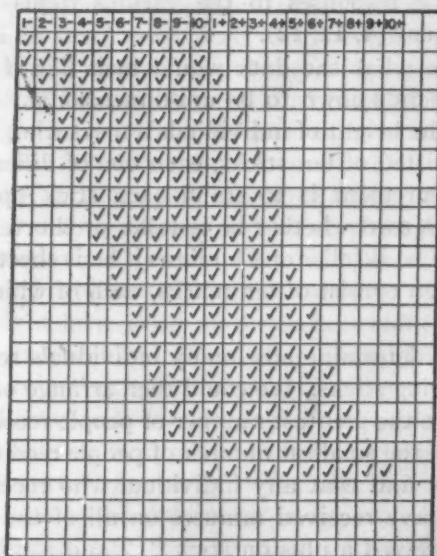


FIG. 1. HYPOTHETICAL EXAMPLE OF A SCALOGRAM

lowing considerations. If hypothetical data from our previous example concerning the ten measuring sticks were arranged in tabular form, each individual representing a row and each column representing a response to a measuring stick, the resultant pattern would be as shown in Fig. 1. It can be shown that for any measuring instrument which is uni-dimensional the rows and columns can be arranged so as to obtain such a "parallelogram." If the measuring instrument does not possess uni-dimensionality, then there will be deviations from this "parallelogram."

The scalogram board is constructed so that any row or any column can be lifted and moved to another position. The procedure of analysis is then to indicate in the appropriate column and appropriate row of the board whether the individual did or did not make that particular response. Then, by shifting rows and columns one can attempt to arrange the board so as to approach as closely as possible to the ideal "parallelogram." If the data can be made to form a parallelogram one can conclude that the measuring instrument possesses unidimensionality. The scale patterns conforming to the various scale positions become apparent from the resulting diagram.

The number of scale positions is, of course, determined by the number of questions in the measuring instrument and the number of possible responses to each question. In the present example where there are ten questions each of which has two categories of response there would be 11 possible scale positions. In general, the number of possible scale types is given as follows (10): "Add unity to the total number of categories in all questions and subtract the number of questions." In other words, if our measuring instrument were made up of seven questions, one of which had two categories, four of which had three categories, and two of which had four categories, the number of possible scale types would be sixteen.

The scalogram board has been used extensively by the Information and Education Division of the War Department. There are several limitations, however, to its use.

1. A scalogram board is fairly expensive to construct.
2. The number of rows in the board must be large enough to accommodate the number of cases in the sample with which the investigator is dealing. If a board is built which will handle 100 cases it will not be adaptable to any sample larger than this. The limitation on the number of questions and number of categories in each question because of the number of columns in the scalogram board is of the same nature.
3. The method is not very rigorous. Outside of possible hunches which the investigator may have concerning how the questions will scale themselves, the method is largely one of trial and error and inspection.
4. Since by the very nature of the method, the investigator manipulates chance deviations to his advantage in trying to form a "parallelogram," an evaluation of the resultant scale is somewhat difficult to make.

*The Tabulation Technique.* The use of this method for "scale analysis" is described by Goodenough (2). It is probably the simplest method available. The steps are as follows:

1. One computes the number of people in the sample making each response to each question.

2. Each question is then graphically represented by a horizontal bar with proportional areas marked off indicating the number of people making each response. These bars for each question are placed directly under one another. Fig. 2 is a reproduction of such a diagram (10) based upon seven questions concerning the desire of enlisted men to return to full time school.

3. Vertical lines are then drawn through all of the bars at the points separating categories of response in each question. The responses between any pair of these vertical lines indicate one of the "consistent" patterns of response.

4. The investigator then goes back to his data to see how well the response patterns of the individuals in his sample fit these scale types.

This procedure is more rigorous than using a scalogram board. It is apparent, however, that in order to use this procedure it is necessary to know beforehand the directions of the responses to each question with regard to the dimension being measured. This means that in order to use this tabulation technique it is necessary to know, for example, that the four possible responses on Question 6 in Fig. 2 go in the direction indicated in the diagram with respect to desire to return to full time school. This is generally not too serious an obstacle.

The originators of "scale analysis" propose that if one finds that the individuals in a sample do not adequately fit the scale patterns one may then examine the data to see if some of the responses to a question may be combined to make the fit more adequate. Thus, for example, an investigator might find that by combining the categories of "undecided" and "yes" on a given question his data might better approximate the scale patterns. The limitation to this is, of course, that again the investigator takes advantage of chance variation and evaluation becomes somewhat difficult.

*The Cornell Technique.* This technique is so named because it has been used at Cornell University. It has been called the "trial scoring and graphic technique" by Goodenough (2). It is described in detail by Guttman (6). It consists essentially of making a trial rank order of individuals and then examining to what extent the responses show the desired type of consistency according to this rank order. Consequent rearrangements in order and combinations of categories are made so as to approximate better and better to the type of consistency expected from a uni-dimensional scale. The investigator can then evaluate to what extent the responses fit the scale patterns. The technique rests on essentially the same principles as the tabulation technique discussed above. It has relatively little to recommend it as compared to the tabulation technique. It is more cumbersome and somewhat less rigorous. Its advantage would seem to be that it preserves somewhat greater flexibility for manipulating the items.

It may be noted that this procedure has been adapted for use with IBM equipment by Noland (8, 9).

*Least Squares Method.* This is another method which was devised by Guttman (3). We shall do no more than mention its existence here since, as Goodenough (2) says, "This method is too laborious to be usable when one is dealing with more than a few items and categories."

#### CRITERIA FOR DETERMINING THE EXISTENCE OF UNI-DIMENSIONALITY

Whichever of the above procedures is used for constructing a scale, the final problem which the investigator must face is that of making a decision as to whether or not his measuring instrument possesses uni-dimensionality. In other words, how well do his data fit those conditions which would exist if uni-dimensionality were present?

For the purpose of making a judgment concerning the scalability of groups of items, the authors of "scale analysis" have introduced the concept of *reproducibility*. From the statements in the published articles, it is somewhat difficult to understand just what is meant. The following are examples of statements made about it:

Perfect scales are not found in practice. In the past, areas the component items of which were at least 85% reproducible from a rank order have been called scales. Recent work shows that it may be desirable to be more stringent about errors and to restrict the word *scale* to areas the items of which are about 90% reproducible (12).

The degree of approximation to perfection is measured by a *coefficient of reproducibility* which is the empirical relative frequency with which values of the attributes do correspond to intervals of a scale variable. In practice, 85% perfect scales or better have been used as efficient approximations to perfect scales (4).

From these statements one would conclude that the reproducibility would be measured as follows: The responses of an individual which exactly fit one of the scale patterns would be 100% reproducible from his scale score. Responses of an individual which deviated by 1 from one of the scale patterns would be  $100(n-1)/n$  percent reproducible where  $n$  is the number of questions being scaled. Thus, if in a sample of a hundred individuals who had each answered five questions, fifty exactly fitted the scale patterns, forty were 1 off a scale pattern and the rest were 2 off a scale pattern, the reproducibility would be 88%.\*

\* Goodenough (2) tends to give a different impression of what is meant by reproducibility. One gets the impression from this article that what is meant by 85% reproducibility is 85% of the individuals falling exactly in the scale patterns. This is obviously not what is meant by other authors.



In an example of a set of items which are considered to show a high degree of scalability one reads,

There are but 16 possible scale types out of a total of 2,592 possible types. Actually almost 400 types occurred among the 2,000 men. Over one-fourth of the men were perfect scale types, and almost all the rest were one or two responses off a perfect scale type (10).

Guttman (4, 11, 12) has also made the distinction between scales and quasi-scales, the latter being measuring instruments which are less than 85% reproducible. As the subsequent discussion will show, this is entirely an arbitrary distinction. It would, perhaps, be better not to speak of uni-dimensional scales (except in very rare instances) and to content oneself with a description of the extent to which the scale on hand departs from the ideal of uni-dimensionality.

Let us now examine how adequate a criterion 85% or 90% reproducibility is for deciding whether or not only one dimension is present. Let us take an example of five questions, each of which requires a true or false answer. We are to determine whether or not these five questions can be scaled, that is, whether uni-dimensionality is present. To make the example more specific let us suppose that to Question 1, 80% of the sample answered *false*; to Question 2, 60% responded *false*; to Question 3, 50% responded *false*; to Question 4, 40% responded *false*; and to Question 5, 20% responded *false*. The number of possible response patterns to these five questions is 32. Only six of these, however, would be "scale patterns." If *false* indicated the same direction of information or attitude on all questions, these six scale patterns would be (1) FFFFF, (2) FFFFT, (3) FFFTT, (4) FFTTT, (5) FTTTT, (6) TTTTT. If uni-dimensionality exists only these six would appear in the data.

From this information alone, namely, the fact that there are six scale patterns out of 32 possible patterns, one could not very well estimate the probabilities of obtaining these six patterns. Let us calculate the chance probability of occurrence of the six scale patterns (holding marginal totals constant), assuming complete independence among the questions. Since 80% answered *false* on Question 1, let us consider that "F" on Question 1 has a probability of .8 and that "T" on Question 1 has a probability of .2. Let us consider the probabilities of false and true responses on other questions in a similar manner. We then come to the conclusion that by chance 42.2% of the individuals would fit the scale patterns exactly. We also come to the conclusion that 48.4% of the individuals would only be 1 off a scale pattern and 9.4% would be 2 off a scale pattern. The reproducibility that one would



get by chance would be 86%. It is clear then that for this specific example a criterion of 85% reproducibility would be a very poor one to employ.

Let us examine what the chance reproducibility would be if we had nine such True-False items that we were to make into a scale. Let us assume that the proportion of "F" answers to the questions are in order .9, .8, .7, .6, .5, .4, .3, .2, .1. There would now be  $2^9$  or 512 possible response patterns. Only 10 of these would be *scale* patterns. One finds, however, that 18.5% of the individuals would by chance exactly fit a scale pattern and that 42.3% of the individuals would by chance be only 1 off a scale pattern. In fact, one finds that by chance one would expect about 83% reproducibility. Again it is apparent that the criterion of 85% or 90% reproducibility would be a very poor one to apply for determining the presence of uni-dimensionality.

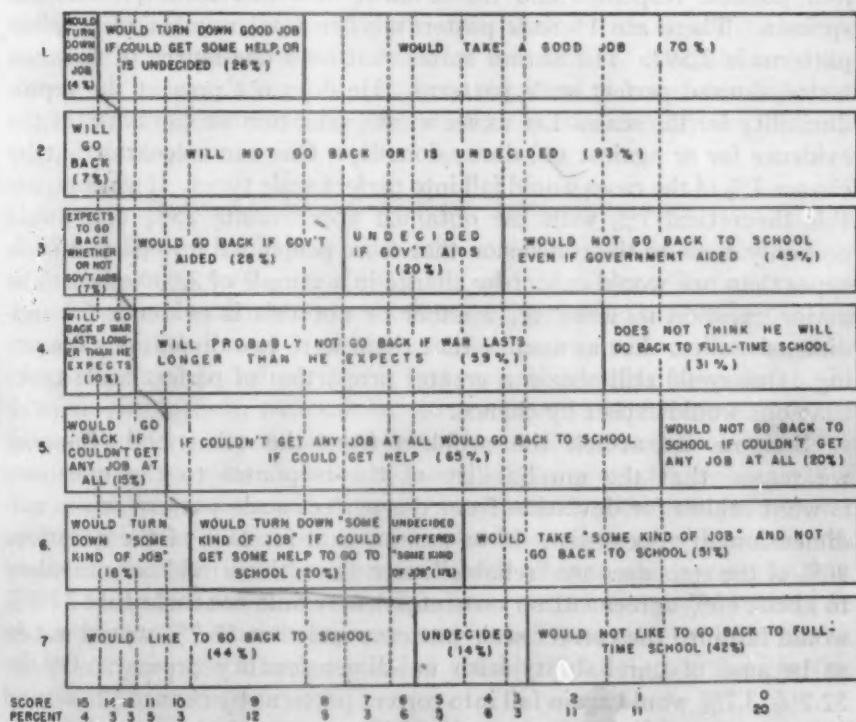


FIG. 2. SCALE OF DESIRE OF ENLISTED MEN TO RETURN TO FULL TIME SCHOOL

It is clear that applying a criterion like 85% or 90% reproducibility to all attempts at scaling, irrespective of the number of items involved or the number of possible answers to each item, leads to false conclusions. In one case where there are many items and many parts to each question, 85% reproducibility might be excellent consistency; in another case 85% reproducibility might represent no better than chance occurrence and be no evidence at all for uni-dimensionality. Any criterion for deciding whether or not there is uni-dimensionality must rest upon calculations based on the number of questions, the number of parts to each question, and the percentages of the sample in each part to each question.

Let us take an example of an actual scale (10, see Fig. 2). Seven items were used. The first item had three possible responses, the second two possible responses, the third had four possible responses, the fourth and fifth items had three possible responses each, the sixth item had four possible responses and the seventh item had three possible responses. There are 16 *scale* patterns. The total number of possible patterns is 2,592. The author states that over one-fourth of the men tested showed perfect scale patterns. He does not present the reproducibility for the scale. Let us see what evaluation we can make of the evidence for or against uni-dimensionality. One can calculate that by chance 7% of the cases would fall into perfect scale types. If we compare this theoretical 7% with the obtained approximate 25% one would certainly come to the conclusion that more people fell into perfect scale types than one would expect by chance in a sample of 2,000 cases. The major question is, however, whether or not this is evidence for uni-dimensionality. Let us assume that there were two dimensions operating. One could still obtain a greater proportion of perfect scale types than one would expect by chance.

Suppose we attack the problem from the other end. Suppose we reason that the unreliability of the responses to the questions is what makes for deviation from the perfect scale pattern when uni-dimensionality does exist. If we assume that for each of the questions 90% of the responses are "reliable" responses, (this would be equivalent to about 80% agreement on test-retest) we would conclude that 47.8% would fall into the correct scale patterns and that 52.2% would not do so because of unreliability with uni-dimensionality present. Of the 52.2%, 3.7% would again fall into correct patterns by chance. The total percentage of individuals that we would expect to find fitting the exact scale patterns would be 51.5%. If this theoretical percentage is compared with the obtained 25%, it becomes clear that, if the assumed

reliability adequately represents the data, more than one dimension is being tapped.

It is suggested here that the expected occurrence of *scale* patterns be calculated assuming (a) independence among questions and (b) uni-dimensionality with a certain degree of unreliability. The obtained percentage of individuals who exactly fit the scale patterns could then be compared with these theoretical percentages using chi square tests. Estimates of the actual reliabilities of the questions used would, of course, make the interpretation unequivocal.

It is the belief of the author that except for the most simple variables, uni-dimensionality will not be found to exist in connection with the measuring instruments which social scientists can construct at present. It would appear futile to insist upon uni-dimensional scales or to make very much of distinctions between scales which possess different "degrees of uni-dimensionality," such as the distinction between scales and quasi-scales.

Scale analysis still provides the investigator with a good technique for scale construction and a means for determining quantitatively the extent to which his data depart from the ideal of uni-dimensionality. Such knowledge should help the investigator considerably in interpreting his data.

#### GENERAL EVALUATION OF "SCALE ANALYSIS"

Thus far the technique of "scale analysis" has been used mostly with the armed forces. Guttman (5), and McNemar (7) both recommend its use in attitude testing and in public opinion research. Limited experience with its use in public opinion research with civilian populations has tended to show that it becomes an unwieldy instrument. In interviewing people, maintaining good rapport and the interest of the interviewee are important considerations. The process of asking as many as ten questions, all of which are, to a large extent, rephrasings of the same thing, is a considerable strain on relations between interviewer and interviewee. Most of those engaged in this type of research will probably find the inclusion of a series of questions which could be subjected to scale analysis not feasible from practical considerations. In connection with standard paper and pencil attitude tests, however, scale analysis offers the promise of considerable improvement in our measuring instruments.

Many claims have been made by the proponents of "scale analysis." These claims are based upon the mathematics of uni-dimensional measuring instrument. Examples of these claims are (4):

Scale scores provide an invariant quantification of the attributes for predicting any outside variable whatsoever.

In imperfect scales, scale analysis picks out deviants or non-scale types for case studies.

We are assured that if a person ranks higher than another person in a sample of items he will rank higher in the universe of items.

These statements may tend to give spurious impressions of the power of the measuring instrument and may create in the user of scale analysis a greater confidence in his instrument than is justified by the conditions of measurement. Non-scale types may occur because actually more than one dimension is being measured or may occur simply because of unreliability of the measuring instrument. Even if a perfect scale were achieved these claims would all be limited by the degree of reliability of the measuring instrument, that is, of the questions asked.

It might be convenient simply to remember that a uni-dimensional scale of very high reliability would have about the same properties as, for example, the common 12-inch ruler. These properties are, of course, quite important, but there is no need to exaggerate them.

If as is suspected uni-dimensional scales do not at present occur even to a good approximation, then these various claims can be mostly ignored as far as the use of the scale is concerned.

It should be emphasized again, however, that the technique of scale construction and evaluation as represented by "scale analysis" seems to be an excellent technique for use with paper and pencil tests or other instances of measurement where the situation permits the inclusion of several questions centering about the same topic. Where used it will provide useful information about the degree of departure from uni-dimensionality.

#### BIBLIOGRAPHY

1. CONRAD, H. S. Some principles of attitude measurement: A reply to "Opinion-attitude methodology." *Psychol. Bull.*, 1946, 43, 570-589.
2. GOODENOUGH, W. H. A technique for scale analysis. *Educ. Psychol. Msmt.* 1944, 4, 179-190.
3. GUTTMAN, L. The quantification of a class of attributes: A theory and method of scale construction. In P. Horst et al., *The prediction of personal adjustment*, Soc. Sci. Res. Council Bull. No. 48, 1941, 319-348.
4. GUTTMAN, L. A basis for scaling qualitative data. *Amer. sociol. Rev.*, 1944, 9, 139-150.
5. GUTTMAN, L. Scale and intensity analysis for attitude, opinion and achievement. Mimeographed, 1945.
6. GUTTMAN, L. The Cornell technique for scale and intensity analysis. Mimeographed, 1946.
7. McNEMAR, Q. Opinion-attitude methodology. *Psychol. Bull.*, 1946, 43, 289-374.
8. NOLAND, E. W. Factors associated with absenteeism in a south-central New York state industry. Doctoral

- Dissertation, Cornell University, 1944.
9. NOLAND, E. W. Worker attitude and industrial absenteeism: a statistical appraisal. *Amer. sociol. Rev.*, 1945, 10, 503-510.
10. *The desire of enlisted men for post-war full-time schooling: an example of a scale.* Research Branch, Information and Education Division, Army Service Forces. Mimeographed, 1943.
11. *Experiments on the measurement of the intensity function and zero point in attitude analysis.* Research Branch, Information and Education Division, Army Service Forces. Report D-1, Mimeographed, 1945.
12. *Questions and answers about scale analysis.* Research Branch, Information and Education Division, Army Service Forces. Report D-2, 1945.



# MANUAL OF CHILD PSYCHOLOGY

A SPECIAL REVIEW\*

ROGER G. BARKER

Clark University

The appearance of the *Manual of Child Psychology* was a major event in psychological publication in 1946. The announcement of a successor to the *Handbook of Child Psychology*, published in 1933 and long out of print, was greeted with enthusiasm, and with the hope that it would prove as useful and influential as its predecessor had been. Many who studied the *Handbook* as students opened the *Manual* with eagerness, for here was a chance to view as a whole the changes in research on child behavior of these 13 years so momentous in science and society.

Some differences between the *Manual* and the *Handbook* are readily apparent as the following comparisons show:

	<i>Manual</i>	<i>Handbook</i>
Approximate length of text	550,000 words	400,000 words
Approximate length of bibliographies	4,400 references	2,800 references
Number of chapters	19	24
New chapters added to the <i>Manual</i> :	Chapters of the <i>Handbook</i> omitted from the <i>Manual</i> :	
<i>Animal Infancy</i> (Cruikshank)	<i>Locomotor and Visual Manual Functions in the First Two Years</i> (Shirley)	
<i>Physical Growth</i> (Thompson)	<i>The Social Behavior of Children</i> (Buehler)	
<i>Environmental Influences on Mental Development</i> (H. E. Jones)	<i>Children's Philosophies</i> (Piaget)	
<i>The Ontogenesis of Behavior</i> (Gesell)	<i>Speech Pathology</i> (Travis)	
and <i>Maturation of Behavior</i> (McGraw) replace Gesell's former chapter on <i>Maturation and Patterning of Behavior</i>	<i>Eidetic Imagery</i> (Kluever)	
	<i>The Physiological Appetites</i> (Blatz)	
	<i>The Child of Special Gifts and Special Deficiencies</i> (Hollingworth)	
	<i>The Child with Difficulties of Adjustment</i> (Blanchard)	
	<i>Birth Order</i> (H. E. Jones) (Incorporated into the new chapter by Jones)	

The quantity of child psychology has increased in these 13 years: in the *Manual* roughly 20 percent fewer topics require 35 percent more

\* CARMICHAEL, LEONARD. (Ed.) *Manual of child psychology*. New York: John Wiley & Sons, 1946. Pp. viii + 1068.

space and refer to over 50 percent more publications. The quality of child psychology has increased too. The *Manual* reports greatly improved techniques and scientific standards, and careful checking and rechecking of 1933 results. In comparison with the *Handbook*, which appears adolescent in retrospect, the *Manual* is a mature, responsible production in the fields it touches.

But a strong impression made by the *Manual* is one of sameness and familiarity. There can be no doubt about it; Child Psychology, 1946 (according to the *Manual*) is surprisingly like Child Psychology, 1933 (according to the *Handbook*). It is true that the old material is more soundly grounded than when reported in the *Handbook*, and perhaps this is all one should expect. Yet these 13 years seemed to be momentous as they were happening. We thought new discoveries were being made, and it seemed that new methods, new ideas, and new attitudes and practices were being developed. One can picture how a similar book in genetics, biochemistry or atomic physics would compare with a 1933 predecessor! We thought Child Psychology, too, was a lusty, growing, rebellious youngster. So we ask, is this report true?

Let us consider how some of the growing points of child psychology are treated in the *Manual*. Take play, for example. Play has become a central technique of clinical child diagnosis and therapy, it is an accepted experimental method, and it has been the object of fruitful investigation. The index of the *Manual* cites play 14 times, but reference to the text reveals that all of these considerations are extremely brief. When we look under the names of those who have contributed to achievement in this area (for example, Bender, Erikson, Lerner, Levy, MacFarland, Murphy, Sliosberg, Stone) we find a similar summary treatment. Here is one of the active areas of child psychology which the *Manual* treats almost incidentally.

We find the same kind of treatment of social behavior. Social behavior has been another productive area: important techniques and concepts have been introduced, and valuable studies made (for example, H. H. Anderson, Criswell, Doll, Horowitz, Isaacs, Jennings, Lippitt, Moreno, Parten, Partridge, Redl, Slavson, Tryon, M. E. Wright, Zeleny). But a student would certainly gain the impression from the *Manual* that social behavior is a moribund aspect of modern child psychology.

Child personality also receives scant treatment. Although Lewin's work is well represented, the contributions of Blos, Davis, Frenkel-Brunswik, Anna Freud, H. E. Jones, MacFarlane, Murphy, Sanford, Symonds, Zachry, for example, are virtually ignored. Personality and

behavior aberrations in children are hardly mentioned. Sigmund Freud is named 15 times, while fifty other contributors are mentioned more frequently.

The interrelations between physique and behavior in children (physiological, psychosomatic, and constitutional) have received rewarding attention in recent years, yet they are given only fragmentary consideration in the *Manual*.

Recent methodological advances have been made in connection with multi-discipline and longitudinal studies, projective tests, play, interview procedures, sociometric techniques, experimental studies of complex behavior and the factor analysis of data. Save for experimental studies, however, although all are mentioned, they fail to stand out from the other methods used with more or less success in the last 40 years. Since the science of any period is a product of its current methods, it is especially unfortunate for the young scholars who may use the *Manual* that modern methodology is not given more prominence.

In the case of theory of child behavior, we find ourselves again confronted with lacunae. One gains the impression that child psychology during the last 13 years has not been concerned with theoretical problems, and that there have been few serious differences in viewpoint. Aside from Lewin's chapter, theory is seldom mentioned. The reader would have to be intuitive to appreciate that a series of battles of viewpoints have been raging across the child psychology front from the school room to the laboratory (differentiation vs. integration, whole vs. part, Gestalt vs. association, discipline vs. freedom, nature vs. nurture, pure case vs. statistical average, directed vs. non-directed therapy, psychoanalysis vs. all comers, training vs. gratification, test vs. clinical impression). Clinical contributions to the theory of children's behavior are not included.

We have to conclude that the *Manual* gives a conservative, almost old-fashioned picture of child psychology by omitting or inadequately reporting some of the most important developments. A careful examination of the text shows that some of these developments are mentioned, but that they are not placed in perspective. Currently important problems and procedures do not stand out; issues are not clearly stated; results are not strictly evaluated. The *Manual* gives the impression of a catalogue, rather than a narrative.

However, these imperfections should not be over-emphasized. The excellences of the volume should be equally stressed. To some, the general conservatism of the *Manual* will be considered its chief virtue. No book can serve all functions or satisfy all tastes. In fact the effort

to do this appears to be one source of the *Manual's* deficiencies. The 1933 *Handbook* could barely accommodate in an omnibus volume the bulk of the material then available and retain some of its variety and individuality; but Child Psychology, 1946, is too extensive to be packed between the covers of this kind of a book without such fractionation and compression that its significance is greatly reduced and its character largely destroyed.

The *Manual* performs in some degree the functions of a number of specialized publications, the functions, namely, of a *reference work*, a *critical research survey*, a *yearbook*, a *bibliography and series of abstracts*, and an *advanced text*. According to the preface, the *Manual* is "an advanced-level textbook," but internal evidence indicates that a number of the contributors had these other kinds of publications in mind to various degrees as well, publications with characteristics that seriously conflict when the attempt is made to include them all in a single volume. Thus, the vast detail which is a virtue in a reference work interferes seriously with the book's usefulness as a text. As a text, the *Manual* requires introductory material which spoils its effectiveness as a critical research survey for mature scholars. The requirements of completeness of coverage and non-evaluation in a bibliography and abstract are at odds with the need for selection, evaluation and organization in a textbook or a research survey. In a book for research scholars, emphasis upon undeveloped but potentially fruitful problems, and unverified but suggestive methods, is a virtue but in a text these are liabilities. The inevitable slowness of producing a book of this size and with this number of contributors seriously interferes with the possibility of reporting current developments.

An additional handicap with which the *Manual* had to contend was wartime restrictions on publication. Because of this its release was apparently greatly delayed; consequently there are not many citations of literature dated later than 1942. For this reason, its immediate revision is desirable. When this is done, it is to be hoped that the *Manual's* special function will be unambiguously defined. We are at the point in psychology where reference works, research surveys, yearbooks, and advanced texts are badly needed, but it is clear from the *Manual* that we are beyond the stage when a single volume can accommodate such diverse materials. These are tasks which some divisions of the American Psychological Association might well undertake for their specialties.

The order of the chapters in the *Manual* and brief comments upon them follow:

*Methods of Child Psychology* (John E. Anderson) is an expanded and

greatly improved version of the same chapter in the *Handbook*. Especially welcome are discussions of sampling problems, design of experiments, actuarial and individual prediction, genotype and phenotype, and statistical and psychological significance. However, this chapter suffers much from the limitations of this kind of publication. Under the necessity of touching upon everything from simple to complex and from new to old, the chapter becomes a conducted tour where much is viewed but where nothing can be examined.

*The Onset and Early Development of Behavior* (Leonard Carmichael) is an expansion of the chapter by the same author in the *Handbook* entitled *Origin and Prenatal Growth of Behavior*. The previous article has become a classic and the present one continues on this same high level. However, one wishes for a book on a subject of this scope; in the present article there are an average of 140 words of comment per research cited. This has been a relatively active field in recent years. The chapter is 20 percent longer and has 43 percent more references than the earlier one; 40 percent of the 500 references are dated 1933 or later. In view of these changes, it is disappointing to find that the issues remain almost exactly where they were in 1933. The author comes to the identical conclusion in 1946 regarding the central issue that he did in 1933, namely "in regard to the related processes of individuation and differentiation of behavior it seems that as yet, at any rate, it is better to record as unambiguously as possible the responses that can be made by a fetus at any stage rather than to attempt to fit all developmental change into one formula."

*Animal Infancy* (Ruth M. Cruikshank) is a valuable addition to the volume. Here is a summary of research in a field that has been tilled little, but which has important potentialities. It is in the finest and most valuable tradition of scholarly research reviews to define and survey such frontier areas.

*The Neonate* (Karl C. Pratt) is an extensive revision and expansion of the same chapter in the *Handbook*. The present contribution is almost twice as long as the original article; 45 percent of the 350 references are dated 1933 or later. While the 1933 review presented a framework, here is the completed structure. On the descriptive level, it appears that the neonate's behavior has been fairly well mapped. It is to be hoped that the frontiers of research will soon be pushed to matters of causation, and from the more physiological to the more psychological problems. The work of Ribble (who is not mentioned in the *Manual*), for example, opens important problems deserving careful investigation. This is another research survey which at times approaches an annotated bibliography (100 words of comment per reference).

*Physical Growth* (Helen Thompson) is a new contribution. To the reviewer's mind, this chapter is out of place. First, as the author states, the material is so vast it cannot possibly be covered adequately. Second,



the crucial problem for psychologists in this connection is the nature of the relation between physique and behavior in childhood, and it is to this that a chapter should be directed. Although this relation has not yet been extensively investigated, it would seem that we have arrived at the point where a short chapter after the manner of *Animal Infancy* is possible.

*The Ontogenesis of Infant Behavior* (Arnold Gesell) is written to quite a different pattern from the other chapters. It is a stimulating essay presenting an interpretation of developmental data in the form of some descriptive generalizations. This essay will be of interest to mature scholars who will argue long and loudly whether these ingenious generalizations are true and whether they are productive of further research and understanding.

*Maturation of Behavior* (Myrtle McGraw) covers the maturation vs. learning problem. Few questions in psychology have had such a turgid course, and here is another competent effort to clarify the issues. Despite the author's protestations of the falsity of the dichotomy, she finds it necessary to wrestle at length with a systematic statement of the relation between learning and maturation. It may be that this was inevitable when the task of writing a chapter with this title was accepted, for it implies that a general solution is possible, which then has to be laboriously disavowed. Perhaps in the next *Manual* the author's conclusion will be heeded and we can have chapters on the *Effects of Training upon Early Behavior Development* and *Neural and Physiological Correlates of Behavior Development in Children*, and save ourselves the pain of struggling with the monster of interdependency again.

*Learning in Children* (Norman L. Munn) is almost twice as extensive as the 1933 chapter by Joseph Peterson. Of the 340 references, 36 percent are dated 1933 or later. In the face of great difficulty, due to the chaotic state of learning studies, the scope of the problem, and the lack of studies directly comparing the learning of children and adults, the author has reviewed a great number of studies of conditioning, acquisition of motor skills, memorizing and problem solving which use children as subjects. He concludes that beyond the very early childhood years when the handicaps of neuromusculature immaturity have been largely overcome, practically all differences in the learning of children and adults may be attributed to differences in motivation and in previous experience.

*The Measurement of Mental Growth in Childhood* (Florence L. Goodenough) is relatively little changed from the *Handbook* version. Of 109 references 34 percent are dated 1933 or later. Additions include the work of Kelley, Thurstone, Thompson, and McNemar on the factor analysis of intelligence test scores, and Bayley on mental growth. This impresses one as being written in the best tradition of an advanced text.

*Language Development in Children* (Dorothea McCarthy) is at the

same time one of the most active and hopeful, yet disappointing, aspects of child psychology reported in the *Manual*. The present chapter is over three times the length of the *Handbook* chapter. Of the 480 references, 40 percent are dated 1933 or later. Here is an active field that touches the basic concerns of child psychology: emotion, personality, intelligence, social development, behavior aberrations, education, neurology. However, few of these larger aspects of language have been subjected to investigation. The problems have been defined within a narrow, technical frame, so that language has been largely divorced from its significance as behavior.

*Environmental Influences on Mental Development* (Harold E. Jones) is a new contribution to the *Manual*. It is packed with information. Here is an adequate reflection of a preoccupation of the times; sixty-two percent of the 235 references are dated 1933 or later. Of all the chapters in the *Manual* this one probably suffers most from the difficulties imposed by the volume. Under the necessity of covering material ranging from I.Q. constancy to diagnosis of identity in twins, and of writing on all levels from pointing out the possibility that coaching affects test scores to discussing the phenomenon of statistical regression to the mean, it has been difficult to present a sharp picture of the issues involved. The chapter will be of great value as a review and commentary for students who bring considerable technical background information, and to those who are able to continue their study in the literature cited.

*The Adolescent* (Wayne Dennis) meets the issue of the *Manual's* ambiguity of function by sharply limiting the scope of material covered. The psychology of adolescence is restricted to the effects of biological adolescence upon behavior. This makes it possible in the space available to give an adequate review of the research upon this problem.

*Research on Primitive Children* (Margaret Mead) is a paper for research scientists. It is a stimulating discussion of the methodology of personality studies in primitive cultures, and one that is almost equally applicable to field studies of children within our own society. This chapter provides a sample of the value of a volume prepared for scholars with no concessions to immature students. A unique and valuable departure is the inclusion in the bibliography of citations of research in progress.

*Character Development in Children—An Objective Approach* (Vernon Jones) is a revision of the chapter by the same author in the *Handbook* entitled *Children's Morals*. This has been a relatively active field; 55 percent of the 140 references are dated 1933 or later. Perhaps because of the nature of the material, this chapter combines better than most the textbook and research-survey functions.

*Emotional Development* (Arthur T. Jersild) is a systematic review of objective studies of emotion in children; 51 percent of the 238 citations are 1933 or later. It appears that the frontiers of this most wild and

difficult region of child psychology are slowly being pushed back. The picture of healthy progress here given is probably an underestimate by reason of two limitations under which the author has worked: 1) the exclusion of all studies not reaching completely to acceptable standards of scientific method and 2) the separation of emotion from other aspects of the behavior of children. It is obvious that some of the potentially most fruitful developments of recent years are as yet outside the bounds of acceptable science and have had to be omitted. The fractionation of the child and the separate consideration of intellect, language, character, learning, emotion, etc. probably is most limiting and misleading in the case of emotion.

*Behavior and Development as a Function of the Total Situation* (Kurt Lewin) is the only chapter in the *Manual* written within a definite conceptual frame. However, it is by no means a speculative discussion. Almost all of the 156 references refer to experimental work; 79 percent of the references are dated 1933 or later. This is the most complete statement of Lewin's position available at the present time. Like so much of Lewin's writing, this paper suffers at points from extreme conciseness.

*The Feeble-Minded Child* (Edgar A. Doll) extends the treatment of the same subject by Pintner in the 1933 *Handbook*. The material has been organized to emphasize standpoints rather than to survey the detailed literature. The topics covered are *definition, classification, incidence, characteristics* and *causation*. Forty-six percent of the 164 references are dated 1933 or later.

*Gifted Children* (Catherine Cox Miles) is in the pattern of the chapter of the same title by Terman and Burks in the *Handbook*. The framework there erected has been filled in with remarkably little alteration. Valuable new material on the later history of Terman's subjects is reviewed. There are 330 references; 42 percent of them are dated 1933 or later.

*Psychological Sex Differences* (Lewis M. Terman and Associates) is a thorough, critical review of the widely dispersed data bearing upon this topic. It will be an important starting point for a long time to come for research scholars concerned with sex differences.

The *Manual of Child Psychology* impresses one with the number of raw facts that have been reliably observed and recorded and with the number of interrelations that have been described. This very wealth of data is becoming a burden. In the absence of a strong framework of theory to organize and subsume such an array of facts, and to guide further observation and experimentation, we are in danger of being engulfed. As a theoretical structure is developed we can expect the definition of child psychology to change from one based on chronology (that part of general psychology which uses chronologically young individuals

as subjects) to one based upon psychological criteria. When this occurs, child psychology will become less disorganized and less encyclopedic and the task undertaken by the *Manual* will be less formidable. This is the basic difficulty confronting the *Manual*: lack of concepts and hypotheses appropriate to the empirical content of child psychology. As a compendium of a great amount of this content, and as a publication base in terms of which more specialized books may be produced, the *Manual* will have a great and permanent value.

## RESPONSE TO CRESPI'S REJOINDER AND CONRAD'S REPLY TO APPRAISAL OF OPINION-ATTITUDE METHODOLOGY

QUINN McNEMAR

Stanford University

The rejoinder of Crespi (4) and the reply of Conrad (2) to our appraisal of opinion-attitude methodology (9) necessitate a published response to a few points, particularly those having to do with facts. We will first consider three of the points raised by Crespi.

1. We acknowledge a serious error in reporting Cantril's material on the reliability of the polling procedure. Though we correctly reported the only figure for the reliability of opinion questions, we did confuse the data regarding factual-type questions.

2. It is said (4, p. 563) that we "overlooked" at least three articles on reliability, and that if we "had seen" these we "would have appreciated the injustice" of our remark that in opinion research individual "reliability has been *practically* ignored by the users of the *single* question technique" (9, p. 314; our present italics). We did, indeed, fail to report on the three cited papers, but not because they were unseen. We concur with Crespi when he says, "Interested readers are invited to check the facts here for themselves" (4, p. 563). What are those facts, and are they pertinent? Jenkins' paper (7) is concerned solely with questions of the type "What brand did you last purchase?," not with opinion questions. Section D of Dodd's report (5) contains empirical findings on reliability—of questions pertaining to radio listening habits—with no specification as between factual and opinion questions. Now it happens that we had also seen the 148 page, security-marked "confidential," document referred to by Dodd (5, p. 266). This does contain information on the reliability of opinion type questions, but unfortunately the data are not analyzed and presented in a manner which makes it possible to ascertain *how* reliable the questions were. King's article (8) deals with 15 questions which are "idea-centered," hence atypical as regards ordinary single question procedures. Furthermore, detail is lacking for evaluating his report that the agreement between results secured from 25 cases by two interviewers, with a two week lapse between interviews, ranged from 60% to 100%.

3. Crespi says (4, p. 567) that our handling of validity is "plain contradiction," which results in "Three errors—an error of fact, an error of logic, and an error of unjustified condemnation." In proving all this, he quotes from p. 315 of our appraisal. At the risk of tiring the reader, we present the quotation along with parts (here bracketed) omitted by Crespi:



One might have expected that Cantril's volume (1) would have tackled the problem of opinion question validity but such is not the case. [Evidently "the serious problems encountered in every phase of the polling operation" (1, p. viii) do not include what the writer considers to be foremost in importance, namely, validity and reliability for the individual response.] The volume does touch on some of the factors which contribute to [unreliability and] invalidity [but a frontal attack is sorely needed].

In what way is any of the above "an error of fact"? We urge the reader to search Cantril's volume for a validity study. Then having found none, he will see that the "contradiction" and the "error of logic" consist merely in having reported, in a credit line, that some of Cantril's material has a bearing on the problem of validity. Crespi correctly gives the *location* of this material when he states that "the entire point of chapter 1 and, in part, chapter 2 of Cantril's volume is the validity of questions" (4, p. 567). Actually, chapter 1 on the "Meaning of Questions" contains no mention of validity, and in chapter 2 on the "Word-ing of Questions" validity is mentioned only in the discussion of the "absence of any objective criteria of validity" (1, p. 23).

We now turn to five specific points in Conrad's reply and then we shall discuss briefly a method used in his evaluation.

1. Regarding the question of item selection, i.e., the *aposteriori* singling out of particular items for study, we too are "at a loss to understand" (2, p. 581) how we overlooked an entire paragraph which not only definitely negates our statement that no reason was given for choosing items, but also supplies information in line with our surmise about possible capitalization on chance. We hereby tender our apology to Drs. Sanford and Conrad for this oversight.

2. It is evident that Conrad does not (and need not) agree with us on issues having to do with scales, uni-dimensionality, and reliability, but when he defends his scales by claiming that the average inter-item correlations of .08, .12 and .07 for his war-optimism scales are "as high as [those] among items of acceptable intelligence tests" (2, p. 572, with repeats of the same idea on pp. 575 and 586), we wonder what empirical data on acceptable intelligence tests he had in mind. He states, without giving a source, that "intelligence-test items ordinarily correlate with the *total* test only about .30-.50" (2, p. 572). Now to some figures. The correlations for items vs. total score for the 1937 Stanford-Binet Revision (10, pp. 176-185) range from .27 to .91, with 189 of 243 coefficients at or above Conrad's upper value of .50. The mean of these 243 correlations is .61. In the case of the Terman-McNemar Test of Mental Ability the average item-total score correlation is .53 for *r*'s based on 1200 subjects (13, p. 2). As to inter-item correlations, the average value for the 1937 Stanford-Binet Revision (10, p. 156) is about .38, and from the above mean of .53 it can be inferred by use of Richardson's Formula 7

(11, p. 70) that the average intercorrelation among the 324 items of the Terman-McNemar test is near .28.

One other point on uni-dimensionality: we agree with Conrad's argument (2, pp. 572-573) that for prediction purposes it is permissible to combine in an index anything that improves prediction, but we were not discussing the problem of prediction. We had questioned the meaning of total scores based upon items of near zero intercorrelation and had criticized as meaningless an index of "life satisfaction" based on four uncorrelated components. Since no outside criteria were available, the problem in both these cases was that of scaling not prediction; hence Conrad has confused the issue by bringing in an irrelevant argument. The reader may refer to Guttman (6, p. 148) for a clear-cut statement of the necessity of keeping in mind the distinction between these two closely related topics.

3. Conrad thinks we are "seriously in error" (2, p. 577) in stating that when an item has a reliability of .25 it can be said that 75% of the obtained response variance is due to measurement errors. Now if we understand Conrad's argument against this well-known statistical fact, he is defining an individual's *current* attitude in terms of the obtained response as though the response were not subject to error. He says that "Current attitudes, thus measured, can be denied validity only if it be assumed that item-responses are indeed evanescent" (2, p. 577). One need not assume evanescence if Conrad's reliability estimate of .25 is correct—such low reliability conclusively proves the assumption.

4. Regarding part-whole correlation, Conrad says that "McNemar is mistaken" (2, p. 578). This is proven by a mathematical derivation which begins with the supposition that "if the set of 10 items had a higher standard deviation than the set of 14." This supposition is contrary to the facts: the 10-item scale had an S.D. of 4.1 and the 14-item scale had an S.D. of 5.9 (3, p. 293). Consequently, if our original contention needed an algebraic proof it has now been supplied by Conrad.

5. Conrad (2, pp. 581-583) counters our criticism of his failure to apply statistical tests by saying that good hypotheses can easily be buried by the mechanical application of significance tests. We agree, but we fail to see how that argument justifies ignoring entirely such tests, especially when doing so leads to the conclusion, concerning "one of the most discriminating questions," that a relationship exists which (when computed by us) turns out to be a mere correlation of .06 for 88 cases. In further defense, Conrad states that "The reader will recognize the 'clinical,' 'hypothesis-hunting' nature of the remarks" which describe briefly the procedure which he and Sanford used. This "hypothesis-hunting" consists of beginning with the hypothesis that two variables are correlated, collecting data, making a scatter diagram (from which it should be immediately obvious that the hypothesis is not borne out), next examining the extremes and finding that the median Y-value for

the high and low X-groups "are the same," then looking at the extremes again and noting that the frequencies in the corners of the scattergram are actually 3 and 0 for the high as opposed to 1 and 2 for the low group, and finally basing the conclusion on 6 of the original 88 cases. Any reader who thinks the foregoing an exaggeration is referred to the printed report (12, pp. 13-14).

6. Aside from the above five factual-type points, we would like to mention a few factors that tend to make Conrad's reply sound like a convincing rebuttal to our appraisal. Repeated reading of his discussion along with checking back to our original paper led to some interesting observations. We noted the frequent occurrence of the debater's "my opponent himself has said," along with the stratagem of giving partial quotations and/or ignoring the context from which quotations were lifted. It is not for us to say whether the use of such forensic devices should be discouraged in scientific writing, but in the present controversy it is not only our right but also our duty to point out specific instances which are apt to mislead the reader who does not have time to do detective work.

a. In connection with the problem of the comparison of dispersions on single items, Conrad's discussion (2, pp. 576-578) ends with "We are inclined, rather, to agree with McNemar's statement at another point in his review, that—," and then he gives the following quotation *without* the part which we here bracket: "[Another advantage of scaled opinions is the fact that the] variation within groups indicates the relative homogeneity of groups in their opinion about an issue" (9, pp. 327-328). The reader, who is led to believe that we had been inconsistent, would never guess that the quotation had been removed from our context of scales to bolster an argument in a context of single items.

b. Another out of context quotation, from which it would appear that at one place we sanctioned that which we disallowed at another, may be found on p. 575 of the reply. Our criticism of the assigning of absoluteness to the so-called neutral point of a scale is countered with "As McNemar himself has said," followed by a quotation from a section in which we were arguing against, not for, *such* usage of scales.

c. At the top of p. 575 a quotation purporting to show our rejection of a certain dictum concerning requisite reliability is followed by another quotation, which is characterized as unimpressive rhetoric and said to be the only reason which we offered for our view. In this case the first quotation stops short of a part so relevant that had the sentence been given in full the reader would have grasped our real reason for questioning the dictum.

d. On pp. 573-574 a quotation and subsequent discussion would lead one to think that we recognized "temporarily" a situation for which our tolerance 61 pages later "seems to have vanished," but the quotation

did not endorse the thing (summing scores on components) being discussed at the later place.

e. On p. 579, "McNemar's own criterion" for a scale is cited as supporting an argument; reference to the original (9, p. 294) indicates that we were not speaking of criteria, but of what attitude measurement permits.

f. The *failure* to consider statistical significance in checking hypotheses (see point 5 above) is supposedly supported in part by a "McNemar himself has said" quotation having to do with the dangers involved in using small samples. We never suggested that these dangers in significance tests could be surmounted by the simple expedient of ignoring the tests.

g. On p. 580, "McNemar has himself objected to" is followed by a quotation taken from a paragraph in which we thought we were merely listing some conclusions, not objections.

h. On p. 585, Conrad says "The following quotation illustrates the unsympathetic 'slanting' of McNemar's account." Then a passage from one of several closely related paragraphs (9, pp. 334-336) is given. By starting and stopping at the right places, the illustration possesses a glimmer of plausibility, and from the subsequent discussion of the passage the reader is supposed to see that "McNemar is wrong when he implies that Cantril has erred." Before accepting either of Conrad's verdicts, the reader will do well to turn to the context to see what was not quoted.

We have not herein dealt with all the things which seem to annoy Crespi, nor have we bothered with many points on which we disagree with Crespi and Conrad. The preparation of this response would have been greatly facilitated if it had been necessary only that we admit our errors.

#### BIBLIOGRAPHY

1. CANTRIL, H. *Gauging public opinion*. Princeton: Princeton Univ. Press, 1944.
2. CONRAD, H. S. Some principles of attitude-measurement: a reply to "Opinion-attitude methodology." *Psychol. Bull.*, 1946, 43, 570-589.
3. CONRAD, H. S., & SANFORD, R. N. Scales for the measurement of war-optimism: I. Military optimism; II. Optimism on consequences of the war. *J. Psychol.*, 1943, 16, 285-311.
4. CRESPI, L. P. "Opinion-attitude methodology" and the polls—a rejoinder. *Psychol. Bull.*, 1946, 43, 562-569.
5. DODD, S. C. On reliability in polling, a sociometric study of errors of polling in war zones. *Sociometry*, 1944, 7, 265-282.
6. GUTTMAN, L. A basis for scaling qualitative data. *Amer. sociol. Rev.*, 1944, 9, 139-150.
7. JENKINS, J. G. Dependability of psychological brand barometers. I. The problem of reliability. *J. appl.*

- Psychol.*, 1938, 22, 1-7.
8. KING, M. B., JR. Reliability of the idea-centered question in interview schedules. *Amer. sociol. Rev.*, 1944, 9, 57-64.
  9. McNEMAR, Q. Opinion-attitude methodology. *Psychol. Bull.*, 1946, 43, 289-374.
  10. McNEMAR, Q. *The revision of the Stanford-Binet Scale*. Boston: Houghton Mifflin, 1942.
  11. RICHARDSON, M. W. Notes on the rationale of item analysis. *Psychometrika*, 1936, 1, No. 1, 69-76.
  12. SANFORD, R. N., & CONRAD, H. S. Some personality correlates of morale. *J. abnorm. soc. Psychol.*, 1943, 38, 3-20.
  13. Terman, L. M., & McNEMAR, Q. The Terman-McNemar test of mental ability. *Manual of directions*. Yonkers: World Book Co., 1941.



## BOOK REVIEWS

SYMONDS, P. V. *The dynamics of human adjustment*. New York: Appleton-Century, 1946. Pp. xiv+666.

Symonds regards this book as carrying on the Columbia tradition of dynamic psychology initiated by Woodworth and Thorndike. He readily admits, however, that his contribution is "not exactly along the lines of its predecessors." Since his text is frankly and enthusiastically psychoanalytic in orientation, many experimental psychologists will regard this as an understatement.

Dynamic psychology is depicted as a study of the whole individual and of how he adjusts to the situations he confronts—how he derives satisfaction of his inner drives from his environment. Process and cause rather than personality status are regarded as the subject matter of the field. In fact, the individual as a total personality, his over-all development and inner organization fall into the background in this volume as processes are developed. This is shown in the treatment of the ego concept which is not discussed in one section but presented fragmentarily throughout the book as various other Freudian concepts are handled nomothetically.

"... Reason and intellect are dethroned as the principal factors in adjustment" the preface states, and "adjustment [is] primarily a matter of the reactions to frustration and the individual's attempt to avoid anxiety. . . . The use of reason in meeting problems of adjustment is reached, if at all, only in maturity as the result of high intellectual endowment and the capacity to profit by experience." "... adjustment is carried on through the impulses, emotions, and by means of the mechanisms which are described herein." It would be interesting to know the extent to which members of the profession have reached this same conclusion and the effect in such cases of this view upon their theory of education, their courses and research in psychology.

Symonds has not been known as a psychoanalytically oriented writer. Many who are acquainted with him as the author of an excellent organization of the personality measurement literature in *Diagnosing Personality and Conduct* will find this volume surprising. The differences in interests and viewpoint shown by the two contributions may reflect a trend in psychology as well as the capacity of an alert, scholarly psychologist materially to change his approach in middle life.

The author purposes to represent the conclusions of contemporary research and discussion on the problem of human adjustment. He states that as he combed the literature he concluded that the psychoanalytic contributions commanded the field. The basic theory of Freud and of ego psychology are presented with the refinements suggested by later students including Horney, Fromm, Reik, Fenichel, Deutsch, Alexan-

der, Klein, and Isaacs—writers whose work is too little known to students of adjustment theory. The volume is not written primarily from a background of system and concepts of traditional psychology with psychoanalytic literature interwoven, but rather the converse. The chapters on drives, frustration and conflict allude to papers by psychologists. Many of the other chapters such as "Love and Self-Love," "Guilt and Self-Punishment" and "Miscellaneous Mechanisms" contain very few contributions from members of the American Psychological Association. On the whole the experimental studies on personality mechanisms seem ineffectively presented.

In his preface Symonds anticipates criticism on the score that statements are made without apparent evidence, which makes the book seem dogmatic. He emphasizes in rebuttal that when one makes an intimate and detailed study of human life through clinical methods, it is done under conditions by which it is not possible to duplicate data. Moreover, he contends, "... when relationships are verified by repetition of scores of investigators, relationships between observed behavior and the dynamic processes which motivate it gain credence." This presentation may seem unnecessarily weak to the many students who come through introductory courses which emphasize objective methods and attitudes, in which experiments are cited for each generalization made. Too, the students from nonscientific curricula who will use this text need a cautious approach to broad clinical generalizations, which this text lacks. It is regrettable that Symonds did not take a more critical attitude toward some of the findings as he presents them, and at least acquaint the reader with the methods of obtaining these data and with well-known attacks on Freudian theory and technique. Furthermore, the questions of relative reliability and validity of data might have been raised. The student is at a loss to distinguish fact from theory and hypothesis. Too seldom is he told that additional studies relevant to a given concept are sorely needed.

Statements such as this will arouse criticism: "... there is also a part of the ego that says, 'Do' and 'Don't' that is also unconscious" (237). And many will want to know the degree of validity of assertions such as: "[early frustration in oral experience produces] mouthy folks—verbose in expression, persons who like to pull on a big cigar or heavy pipe, or chew on the stub of a pencil."

The mixed style and irregular organization within chapters will irritate many of the more perfectionistic readers. The integrated pattern of psychoanalytic insights tends to be lost as Symonds attempts to organize mechanically the literature in terms of single mechanisms.

Symonds deserves great credit for his efforts toward the integration of a body of widely varied literature, for his attention to truly significant, complex and somewhat inscrutable problems, for his wholesome psychological treatment of social problems arising from individual

maladjustment, and for his over-all organization of the major concepts of dynamic psychology. This respect is enhanced when his volume is compared with some other publications in the field. The 63 pages devoted to notes concerning the 883 references are a contribution in their own right. This work should give pause to the personnel worker who dispenses superficial advice and recommends social maneuvers to alleviate deep conflicts.

FRED MCKINNEY.

*University of Missouri.*

BENEDEK, THERESE. *Insight and personality adjustment: a study of the psychological effects of war.* New York: Ronald Press, 1946. Pp. xii + 307.

As a psychoanalyst of the staff of the Chicago Institute of Psychoanalysis, Dr. Benedek is probably fairly orthodox. As the writer of a book, practical and timely, and geared to the immediate clinical problems of today, Dr. Benedek seems, to the reviewer at least, to differ sharply from most orthodox psychoanalysts. With brief acknowledgment to the theoretical substructure of psychoanalysis, the author addresses herself promptly to that large body of personalities today who reflect the tremendous need for personal, postwar readjustment. The veteran is given careful consideration, not as the *species veteran*, but as he is a husband or a father or a son. What he has in common with other veterans—the war itself—is evaluated psychodynamically as an homogenizing influence. Just as carefully does the author evaluate the war as an experience for the wife and the mother of the veteran, the father and the child, the brother and the sister.

To infer from this that the author is focussed on personalities would be unfortunate, for her concern seems to be primarily with situations: separation, continued separation, reunion, personal readjustment. These situations and the behavior they demand are presented each as a psychodynamic equation.

To the persons involved with it in various ways, the war experience served or failed to serve needs, and an analysis of it in these terms is a healthy exercise for those who today must consider problems of readjustment. Dr. Benedek states her hope that this book will be found useful by social workers, clergymen, teachers, counselors, as well as psychologists. That many such readers will be stimulated by her book I have no doubt, for I liked it. It is clear that Dr. Benedek has thought her way through the present world crisis as it is involved in clinical work. Much more than that, she reveals an insight into present-day problems of personal adjustment which must derive from more than psychodynamic theory and thinking alone. She has seen real people trying desperately to meet today's problems.

Just how this book will be used by counselors, it is difficult to guess;

for it is in a sense a compendium of dynamics, codified according to the interpersonal relationship involved. Whether or not the book will serve as such a reference work, the reading of it as a whole will provide an intelligent psychodynamic analysis of the present world crisis as it troubles the individual today. As a book about individuals this is a book in clinical psychology, and I think it has, from this viewpoint, an unfortunate title.

THOMAS W. RICHARDS.

*Fels Research Institute.*

HEWITT, L. E., & JENKINS, R. L. *Fundamental patterns of maladjustment; the dynamics of their origin.* Institute for Child Guidance, State of Illinois, 1946. Pp. 110.

From the files of the Michigan Child Guidance Institute 500 case records were selected and certain information in these records entered on a code schedule. "Three hypothetical behavior syndromes were suggested including the following items: (1) assaultive tendencies, temper displays . . . etc.; (2) gang activities, . . . stealing, . . . truancy, etc.; (3) sensitiveness, seclusiveness, shyness, etc." These were formalized as (a) unsocialized aggressive behavior, (b) socialized delinquency behavior, and (c) over-inhibited behavior syndrome.

A very large number of tetrachoric and biserial correlation coefficients were obtained between the items and combinations of items. The mass of the statistical analysis is presented in a fashion which the reviewer found confusing. He was unable to see the meaningful forest because of the crowded numerical trees. It is his impression that most of the traits (items) which intercorrelate highly, for example, .67 between seclusiveness and shyness, are self-evident. No clear line of evidence was apparent indicating that other than obvious relationships were found. The evidence may be there, but if so it is obscured by too many statistical tables.

Sample case histories are given to illustrate the personality development of the three syndromes. Dr. Jenkins adds a chapter giving the psychiatric interpretation of this study, together with its implications in guidance and treatment.

The authors concluded that,

This study has served to provide a statistical verification of certain previously established hypotheses . . . concerning the nature and backgrounds of the aggressively delinquent and the pseudo-social delinquent. The situational correlates of the third type of behavior (overinhibited behavior) have been revealed through empirical analysis. In each of the three behavior-situation pattern relationships there appears to be some evidence that not only is the behavior in question provoked by a particular type of frustration but the general pattern of behavior itself is exemplified by other persons with whom the child is in



close contact. Thus the resulting type of maladjustment would appear to be a rational reaction of the child to his distorted environment in a double fashion.

C. LANDIS.

*New York Psychiatric Institute  
and Columbia University.*

THORPE, LOUIS P. *Child psychology and development*. New York: Ronald Press, 1946. Pp. xxvi+781.

This book is intended as a text for courses in education and psychology. Many instructors, however, will find it more useful as a reference source for their own use, and that of their more advanced students. The book is long, repetitious, and at times the writing is not clear. In some instances the selection of research materials, and in many cases their interpretation, is extremely biased. For these reasons it does not meet the needs of the average undergraduate.

Each chapter begins with an historical and theoretical discussion of the topic to be presented. There follows a discussion of findings, including a series of presumed typical investigations and a "summary and implications" in support of Thorpe's own point of view. The book is heavily documented through the use of many long quotations from the educational and psychological literature and numerous footnotes. Questions for study and a list of references are appended to each chapter.

Chapters two, three, and seven, *What the Child Inherits, Mental Abilities: Nature and Nurture, and Intelligence and How It Develops*, deal with the operation of heredity and environment in the process of child development. Thorpe obviously has strong environmental leanings and is, of course, entitled to his point of view. However, this reviewer questions the use of such terms as "congenital inheritance" (pp. 71-72) and "experiential maturation" (p. 144). Few psychologists today would deny the mutual interdependence of hereditary and environmental influences from the moment of conception. But it seems unnecessarily confusing to use "inherited" for such environmentally produced effects upon the child while *in utero* as those caused by malnutrition, disease, or injury to the mother, even with "congenital" as a qualifying adjective. Nor does it seem justifiable to nullify the term "maturation" which by definition means that part of the developmental process which is impelled from within, by preceding it with a word descriptive of extrinsic factors. Thorpe draws a clear-cut distinction between physical attributes which he concedes to be inherited, and all mental and personality traits which in his opinion have neither structural nor innate bases.

Although both sides of the nature-nurture controversy are presented, researches supporting the hereditarian view are sometimes out-dated



where more recent and better investigations are available. The usual misconceptions by which the controversy is perpetuated are reiterated in this book: accepting test results at face value, after warning others not to do so, assuming that because certain conditions invalidate the use of tests that the function tested has changed rather than that the measuring instrument has been rendered useless, and finally, in lieu of conclusive evidence, holding the view that intelligence per se is improvable because we would like to think that it is so. It is unfortunate that these chapters take up so large a portion of the first part of the book. Many will read no further and miss some of the more acceptable materials presented later on.

Thorpe rejects *instincts* and *drives* in favor of *basic needs* as the underlying motivational forces of child behavior. The child continually strives to satisfy these needs, and society's role is to see that he succeeds in a socially acceptable manner. Since the author recognizes no innate personality limitations, a suitable environment is all that is required to produce children free from frustration and conflict. The individual parent is warned against "over-acceptance" of the child, but apparently there is no limit to the extent to which society-in-general should plan an environment accurately attuned to the child's needs.

Emotions are "stirred-up states of the organism," resulting from failure of the child to satisfy his needs. The traditional (puritanical) point of view, that strong emotions are undesirable and to be avoided, is upheld by the author. Pleasant emotions receive scant attention.

In keeping with Thorpe's strong environmental bent, the best chapters in the book are those dealing with *Effects of Early Home Conditions*, *The Social Education of the Child*, *Safe-Guarding the Child's Personality*, and *Mental Hygiene*. An extensive presentation of theories, terms, and research findings is followed by many sound suggestions by which parents and teachers can improve the conditions under which children live and learn.

For the mature student, capable of making his own selection from the wealth of material presented, the book will be a welcome addition to his sources of reference. He will find useful, also, the numerous tabulated summaries of research studies.

KATHARINE M. MAURER.

*University of Nebraska.*

SCHROEDER, E. M. *On measurement of motor skills; an approach through a statistical analysis of archery scores.* New York: King's Crown Press, 1945. Pp. xvi+210.

An interest in measurement and testing in the field of physical education activities was the motive for the research reported in this book. Archery was selected as the subject of study on the ground that it is a highly routinized activity carried out under relatively well stand-

ardized conditions; its score is objectively measured, accurately recorded, and commonly accepted as a valid measure of the performer's skill.

Based on data collected over a period of six years from the performances of 258 beginning, 139 intermediate, and 53 advanced archery students at Wellesley College, and limited in scope to the investigation of the behavior of the score, the study is divided into two parts. In Part One, which deals with the trend and reliability of the score in successive lessons, the purpose is to evaluate score units of stated length as tests of group and individual ability and improvement. Part Two concerns the behavior of the score within a single lesson with the object of determining whether successive small units of score reflect the effects of practice and fatigue, and whether such a lesson is long enough to measure individual skill accurately.

Because of the standardized nature of archery, many of the findings of the study—such as, for example, that the sum of the best two scores estimated individual skill more accurately than did the sum of the last two scores of the series of lessons, that the Range score on twenty-four shots was sufficiently reliable to be a useful test of group ability, that the lesson of standard length was too short to measure accurately the skill of the majority of subjects—will have direct application to testing in that field. Taken as a whole, however, the results obtained illustrate and emphasize a fact which is of prime importance in the interpretation of test scores, namely, that a score made at any stated time is likely to estimate ability with varying degrees of accuracy. For the light it throws on the nature of scores, the study should be helpful in suggesting procedures to improve the accuracy of measurement in other sports activities.

The methods by which the various types of problems undertaken were handled are presented in detail. Careful definition of terminology and the adequate treatment given to the statistical techniques employed add to the effectiveness of the presentation. As an exposition of testing problems that are unquestionably fundamental, the book is an excellent contribution.

JAMES M. LYNCH.

*Personnel Research Section, AGO.*

EDWARDS, ALLEN E. *Statistical analysis for students in psychology and education*. New York: Rinehart, 1946. Pp. xviii+360.

This book is intended primarily for students of psychology. The conventional chapters on scaling, reliability, and validity are omitted on the grounds that they belong in courses on tests and measurements, while consideration of multiple correlation is postponed to more advanced courses. The author begins with a simple review of basic arithmetical computations, such as fractions, decimals, square roots, posi-

tive or negative numbers, and then goes on to the treatment of measures of central tendency and variability. A notable omission is the graphic treatment of distributions, and this is regrettable since some guidance in the graphic presentation of frequency polygons (or histograms) is an essential part of a student's training in elementary statistics if his training is to help him in subsequent laboratory or field work in psychology.

The treatment of fiducial limits and reliability of means, which so often is either garbled or even erroneously presented in some elementary texts, is treated here accurately but in a rather verbose manner. It is gratifying to see that analysis of variance has been finally divorced from the barn-yard imagery which adheres to this term in agricultural and biological textbooks.

The chapter on  $\chi^2$  is primarily devoted to the treatment of qualitative or enumerative data and it would seem better to give it such a title. In general, this tendency to deal with the names of special techniques as chapter headings rather than with the contents of these techniques is unfortunate, for it gives the student the feeling that statistics is made up of disparate tools. It would be much better to organize the content according to topic and let  $\chi^2$  and comparison of percentages, for example, fall into the same group. Similarly, analysis of variance could well form part of a chapter or a series of chapters on comparison of means and thus give the student an integrated point of view regarding critical ratios,  $t$  tests,  $F$  ratios, and correlation ratios. This is done to some extent on p. 199, but because of the segregation of the techniques into separate chapters it may appear to be an accident rather than a necessary result. By bringing these methods together, the consistency of different treatments can be demonstrated and the unity of statistical method stressed.

The final chapters on sampling prediction and design of experiments fulfill a basic need in an elementary text.

In general, the text's chief advantage lies in the fact that it draws its material entirely from psychological sources and thus makes the student of psychology feel quite at home within its pages. In some respects, it descends to quite an elementary level, presuming very little, perhaps too little, on the part of the student. In other respects, especially in the chapter on analysis of variance, it goes far beyond the proper limits of an elementary text.

In common with the majority of elementary texts, this book devotes little space to the consideration of even simple curvilinear relationships with the result that the student gets the impression that psychological data are primarily linearly related, and that functional or curvilinear relationships do not exist, are due to artifacts or are beyond the pale of psychological investigation. As a consequence, the student of psychol-

ogy is completely at a loss when he begins to deal with such nonlinear relationships as the Weber-Fechner Law, the many functional relationships in the field of visual and auditory perception, and the growth curves in developmental psychology. The treatment of functionally related variables is no more complicated than that of analysis of variance or the correlation coefficient. Since it is so useful to the student of psychology it should find a place in the elementary texts.

JOSEPH ZUBIN.

*Columbia University and  
New York Psychiatric Institute.*

GOTTSCHALK, L., KLUCKHOHN, C., & ANGELL, W. *The use of personal documents in history, anthropology and sociology*. Social Science Research Council Bull. 53. New York: Social Science Research Council, 1945. Pp. xiv+243.

This Bulletin comprises three separate and distinct monographs. Gottschalk gives brief and concise descriptions of historiography and historical method in the use of documents. In similar fashion he deals with the principles of external and internal criticism of documents. These principles are usually more important to historians than to psychologists and social scientists, since the latter seldom have reason to question the authenticity of documents which come to them. Nevertheless, social scientists frequently use historical method, and Gottschalk's suggestions regarding the use of documents should help them to avoid pitfalls in their excursions into historical fields.

Kluckhohn points out that anthropologists, unlike their colleagues in other fields, seldom work with materials which have been produced spontaneously for confidential use. The documents employed by anthropologists are usually written at the instigation of field workers, or are merely records of interviews. Kluckhohn's critical review of such documents, including a number of biographies, is an excellent guide to the literature. The suggestions which he makes for future research, and for making available more of the original and detailed notes from field work, represent no mean contribution to methodology and technique in anthropology.

Angell catalogs and discusses the various ways in which personal documents are used by sociologists, emphasizing that such materials both yield hypotheses and contribute to the verification of hypotheses. He reviews the principal sociological studies of documents under two headings: (1) explanation of historical sequences, (2) contributions to sociological theory. A brief chapter on sociological method is followed by one on suggestions for obtaining, analyzing and interpreting documents and demonstrating the validity of hypotheses from documents.



No brief review can do justice to the importance of these monographs or to their usefulness as guides to users of documents. Nor can they be evaluated properly without considering them in the light of publications of a similar nature by Dollard, Blumer, and G. W. Allport, the latter two of which are authors of previous SSRC Bulletins. The definitions of *personal* document (or *human* document, as the case may be) given by these six writers are so different and distinct that they suggest the analogy of the blind men describing the elephant.

Blumer defines the *human* document as "an account of individual experience which reveals the individual's actions as a human agent and as a participant in social life."

Allport defines the *personal* document as "any self-revealing record that intentionally or unintentionally yields information regarding the structure, dynamics, and functioning of the author's mental life."

Angell, for his purposes, defines a *personal* document as "one which reveals a participant's view of experiences in which he has been involved" (p. 177).

Kluckhohn agrees with Dollard that the subject of a personality sketch or biography "must be viewed as a specimen in a cultural series," but adds that "the analysis will be incomplete unless the universal, communal role and the idiosyncratic components of personality are distinguished" (p. 138).

Gottschalk, in contrasting Blumer's definition of *human* document with Allport's definition of *personal* document, regards these adjectives as tautological, since "every document, no matter how thoroughly the author strove to be objective, must exhibit to a greater or less extent the author's philosophies and emphases, likes and dislikes, and hence betrays the author's inner personality" (p. 13).

These definitions are quoted not so much to emphasize that the representatives of the several academic disciplines are not in complete agreement regarding their fundamental concepts, as to point out that in stating their definitions boldly and precisely these writers have provided the key to an understanding of the distinctive methods and techniques characteristic of sociology, psychology, anthropology, and history applied to documents. In these days when psychologists, social scientists, and historians are being encouraged to pool their concepts and techniques in grand "cooperative projects," there is strong temptation to urge that all of them should get together and agree on fundamental concepts and methods. While such cooperation is desirable, it must be remembered that data, interpretations, and generalizations are in part functions of the concepts, methods, and techniques employed in the research which produced them. Pooling concepts, methods, and techniques may result in data of doubtful value, interpretations devoid of clarity, and generalizations lacking precision and validity, unless the cooperators have scrupulous regard for the various methodologies.



These monographs are excellent guides to such methodologies, at least in so far as the use of documents is concerned.

ARTHUR F. JENNESS.

*Williams College.*

LILLIE, RALPH S. *General biology and philosophy of organism*. Chicago: University of Chicago Press, 1945. Pp. 215.

The question—What is life?—constitutes the central theme of Professor Lillie's book. That he is well qualified to discuss this problem will be denied by few; that he has been successful in his attempt to formulate an answer will be denied by many.

It will come as no surprise to those who have read Lillie's many contributions to the *Journal of Philosophy*, *Philosophy of Science*, and other scholarly journals, that he endorses a teleological interpretation of life. Such a vitalistic view is somewhat novel in a day when scientific sophistication appears to be synonymous with an acceptance of physical determinism.

Following a brief review of cellular chemistry, Lillie attempts to show that the basic fact of chemistry is that the orderly and complex distribution of molecules requires energy. He feels that this energy must be applied "—directively within the organism," in order to overcome the natural tendency for molecules to become distributed into random and symmetrical patterns, a condition which is presumed to be due to chance. This natural tendency is expressed in the second law of thermodynamics and is a tendency which must be overcome if there is to occur any cellular differentiation. The required energy, which he refers to as an anti-diffusion factor, is, according to Lillie, essentially teleological. That is, it is applied directly with some purpose.

For him, the organism is a psychophysical system in which the final integration of the physical and psychical factors appears to be psychical rather than physical. The physical is characterized by constancy, by permanence, and by a static nature, whereas the psychical is novelty, is change, is process and occurs only in the present. The psychical cannot be observed externally as can be the physical but is instead felt and self-experienced. This is characteristic of one of the concepts of phenomenology. He expressly disclaims a dualistic view but rather feels that the two fields, the physical and the psychical, are grounded in the same fundamental reality. This corresponds closely to the monistic double-aspect view of psychology.

According to Lillie, the entire psychophysical organism acts as if it were the field of a specifically unifying factor, a factor which has constant properties but whose activity is directive and synthetic in essence. This action is carried out in pursuit of definite purpose and "psychic aim" becomes dominant and determines the special direction of what happens in the physical world. For example, Lillie rejects the theory

that the natural selection of purely fortuitous variations (*e.g.* mutations) is a sufficient explanation of the origin of the entire range of adaptive characters. He definitely feels it necessary to explain evolution in terms of a directive purpose.

The crucial point in Lillie's theory, and one which he quite clearly recognizes as being of vital importance, is how the psychical can operate on and give direction to a physical system when the psychical factor, as such, cannot exert physical force. He attempts to get around this dilemma by placing the locus of psychical control internal to or behind the elementary physical events (quantum transfers). How this solves the dilemma is not clear. It appears to be a restatement of the problem in another way and on another level. While some may feel satisfied that Lillie has found a locus for the psychical factor (the reviewer does not), the way by which it can exert physical force is still an enigma.

How then are we to evaluate this book? Is it possible that the author has stated his views too soon, that he should have waited for more evidence? If Professor Lillie were at the outset of his career, we might reasonably expect him to remain an agnostic until more evidence should become available. But he is not at the beginning of his career. He is already an emeritus professor. Furthermore, his lifetime in the laboratory has given him much experience that could be obtained in no other way. And is it not worthwhile for younger scientists to learn of his experience?

It is probably impossible to correctly say whether Lillie is right or wrong in his assumption of a teleological interpretation of nature. True, he does not have the weight of numbers in support of his position, but for that matter neither has any other scientist who has dared to break with tradition. Surely the history of science has taught us the folly of mere majority opinion. It must also be taken into consideration that there is more hope in Lillie's view than in that of the physical determinist, but, unfortunately, wishing does not necessarily make a thing so. Perhaps the chief merit of the work is in the possibility that it will stimulate further research on the problem from which an unequivocal solution may be derived.

ROBERT P. FISCHER.

*University of Florida.*

FALES, WALTER. *Wisdom and responsibility—an essay on the motivation of thought and action.* Princeton: Princeton Univ. Press, 1946. Pp. 166.

In this closely-written, highly epigrammatic, and occasionally obscure series of ten essays, an objective theory of value-experiences is presented. Like Koffka, the author holds that we are earlier aware of the subjective values of objects than of their spatial and temporal attributes. These "subjective values" are *motives*; objective values are

mental categories in terms of which motives are "innately" interpreted and judged. Intentions are prior to thoughts, decisions prior to insights. We are able to organize because we are ourselves previously organized and to create because we are creatures. "Everything which has weight in a man's life or meaning in his thinking derives its structure from his final ends" (p. 161).

Individuality itself is nothing but the ability to regulate the growth of motives in such a way that situations become solvable (p. 26). Every individual develops in a chain of decisions which can be understood as evidencing a system of objective values; but our genuine decisions are determined by final ends which we do not even see although we have a relation to them. Personality or "model individuality" is the response to *calls* rather than to *needs*, i.e., we stand for something that is bigger than we are (p. 68).

Educationally, this outlook leads to the important suggestion that it is the function of intuitions to perceive wholes which account for the coherence and meaning of that which otherwise would remain below the threshold of interest and attention. For example, knowledge obtained by way of inference is encyclopaedic and the property of all who are intelligent enough to take possession of it; but more fundamentally, those impulses that lift us up to specific planes of understanding are symptoms of an anxiety which drives us to master the problem "world" in accordance with standards set by our final ends (p. 148). This defines the problem of the educator as the control of the momentum of powers active in shaping the better selves, i.e., the system of demands of other persons—all because human life is essentially obligation.

Superficially, this seems like a contemporary restatement of ethical transcendentalism, but on closer scrutiny it takes on more of the character of a novel form of radical empiricism or field theory that seeks to place the Jamesian marginal in experience on the same or even loftier ontological footing than the focal regions of our phenomenal world. Only in such a context does the assertion that we are least selfish when we are most creative become intelligible. The key suggestion that motivation is ultimately impersonal in the sense that the organism relays forces greater than itself is truly profound, and may compel some useful revisions in the need schemata now popular among applied psychologists and "social engineers."

GEORGE W. HARTMANN.

*Teachers College, Columbia University.*

SARGENT, W. E. *Teach yourself psychology*. Philadelphia: David McKay, 1946. Pp. 159.

*Teach Yourself Psychology* is the sort of popular presentation that one hopes will not have a wide audience. The first chapter, consisting of little more than a listing of the names and dates important in the pre-

experimental history of psychology, furnishes some basis for believing that this hope will be justified. In the somewhat more readable discussions that follow, covering among others the topics of instincts, structure of the mind, and dreams, most of traditional experimental psychology is included in a single chapter. This reflects, of course, the author's clinical (and religious) bias. He is still fighting the battle against Watsonian behaviourism, and he fights with the same religious zeal manifested by critics 20 years earlier. His own theory is taken in about equal parts from Freud and McDougall. The tone of the volume is typified by the following sentence: "Further, if man is nothing more than a machine which reacts to the world around him as a typewriter responds to the touch of the typist, how can we explain the visions of prophets and the dreams of seers whose thoughts are unique and in advance of their generation." The concluding sentence contains another example of the same sort: "Its (psychology's) final aim is not merely to state how man thinks, feels, and acts, but how he can do these things much better and more in accordance with the Divine purpose that lies at the back of all things; . . ." If the present volume is typical of the rest of the Teach Yourself series, the publishing effort involved would seem to be quite ill-advised. There is little question, however, that it is easier to make a mistake of this sort in psychology, than, for example, in algebra.

LLOYD G. HUMPHREYS.

*University of Washington.*

BLUMENFELD, WALTER. *Introducción a la psicología experimental*. Lima: Editorial Cultura Antártica, S. A., 1946. Pp. 417.

For more than a decade Dr. Blumenfeld has furthered the interests of psychology in Peru, and the influence of his prolific and constructive activity has extended over much of South America. His production of this introduction to experimental psychology in Spanish is likely to be of great importance in establishing psychology as an autonomous field in regions where it exists largely in an incidental way in relation to other interests.

The book is a *précis* of experimental psychology, covering in 24 well-arranged chapters all the topics one would expect, supported by the inclusion of very recent data. A full bibliography is drawn from a wide range of experimental literature and convenient tables, clearly printed charts, and excellent diagrams including some plans of laboratory apparatus, are extensively used. In style of language the exposition is consistently clear and economical. Publication was sponsored by the Instituto Psicopedagógico Nacional, which organization, along with the publishing concern, deserves considerable praise for a well-executed undertaking. In a straightforward preface the author tries to make clear to his readers the basic importance of experimental procedures for all psychological work, and thus provides an emphasis that has been



lacking in some parts of Latin America. In discussing such topics as psychoanalysis and personality he shows an enlightened scientific regard for the bearing they have upon experimental psychology. More than half the book is devoted to sensory and perceptual processes, and the psychophysical methods are well presented in their historical contexts as well as in present applications. It is interesting to observe through Dr. Blumenfeld's portrayal the greatly varied utility of rather a small number of scientific principles for building up the total experimental field.

Without introducing any dogmatism of theoretical approach the author quietly and effectively achieves his intention of allowing experimental psychology to speak for itself. The logical background is highly eclectic. And when it becomes necessary to make a break with the somewhat speculative philosophical tradition of Latin-American psychological thinking, the break is always made gently, considerately, and courteously. The strong presentation of a needed emphasis amidst an atmosphere devoid of polemics is a tribute to Dr. Blumenfeld's grounding as a psychologist and skill as a writer.

HOWARD DAVIS SPOERL.

*American International College.*

LINK, HENRY C., & HOFF, HARRY ARTHUR. *People and books*. New York: Book Manufacturer's Institute, 1946. Pp. 166.

Will the war-time boom in book-buying decrease with the peacetime availability of essential and luxury items? The true guide to the future of books is the reading habits of people, not just the present sales figures. In 1945 the book industry's business reached a half billion dollar total and yet up until that time no one had scientifically investigated the market.

Through the combined efforts of all the major groups of businesses which participate in the production of books, the Psychological Corporation and the Hopf Institute of Management were retained to meet the need for a reliable and accurate index of consumer book reading and buying habits. An extensive survey was organized to answer the main questions about which the publishers and others of the book industry had heretofore been able only to conjecture and theorize.

Four thousand interviews were conducted with consumers between May 21 and June 8, 1945, after a six month period of developing methods and revising questionnaires. Acting as field supervisors, sixty-two psychologists directed the work of the interviewers in 106 cities and towns.

The results of this research show clearly that income is *not* the dominating influence, but rather that formal education is the determining factor in book reading. To be sure, people have price preferences in books, but these are not based entirely on economy; often they reflect



motives which can be appealed to by the makers of both the lower-priced and the higher-priced books.

Book stores, book clubs, and department stores, in that order, are the principal sources for the consumer purchase of books. Purchasers do not mean readers, however. In this study it was found that best-seller lists give no indication of the number of people actually reading the books. During the period preceding this survey the reading of *Forever Amber* equalled if not surpassed the reading of the Bible!

The results show that the population spends twelve times as much time per day on newspapers, magazines, radio and movies as it does on books. The reviewers do not feel, however, as did the investigators, that this necessarily indicates so great an amount of *available* book reading time. We feel that before any such generalization as to potential book markets can be made, the *purposes* for which newspapers and magazines are read and the activities-while-listening of radio users must be ascertained.

Forty-one per cent of the persons surveyed in this study said they own less than 100 books; thirty-four per cent claim to have more than 100 books. The accuracy of respondents estimates was not empirically verified by a sample.

A parallel survey was conducted with dealers, distributors, publishers, and educators. The results corroborated the consumer survey in predicting an expanding market for books.

Two suggestions occur to the reviewers for possible improvement of the study: (1) an analysis of conditions under which the book industry should decide to contract, maintain, or expand future production facilities, so as to utilize more fully the present data, and (2) the development of a predictive formula for future commitments to be based upon the comparison of book sales with survey findings when, as the investigators suggest, surveys of this type are repeated.

NANCY C. COOLEY.

ROBERT H. SEASHORE.

*Northwestern University.*

CLEETON, G. U., & MASON, C. W. *Executive ability: its discovery and development.* (2nd Ed. Rev.) Yellow Springs, Ohio: The Antioch Press, 1946. Pp. iii + 540.

According to the authors the purpose of this book is to report and coordinate the best available information concerning the qualities necessary for proper performance in executive positions. The topics treated range from neuromuscular activity to a definition of democracy. Emphasis is placed on problems of selection and training and the role of executives in labor relations. The treatment appears to be strongest in the last area and less strong when the discussion turns to scientific principles of behavior. Typical of the latter case are the acceptance and use

of terms such as instinct, will, and mental capacity without a clear statement of operations defining them.

The major shortcoming of this volume is that sweeping generalizations, presumably based on research findings, are reported without the necessary information to evaluate the instruments, procedures, statistical tools, and data. A few other criticisms may be listed. The presentation of selection procedures (roughly 125 pages) omits the critical concept of cross validation. Over-emphasis on rating scales leads to the ignoring of more objective measures of proficiency. The authors believe that there is no communality among the traits necessary for success in different executive positions; this position implies a research program for each executive job, certainly an impractical task.

It is the opinion of the reviewer that the attempt to treat this important applied field is admirable and that this book may be useful in presenting elementary notions for the non-psychologist working in this area. It is felt, however, that the book falls short of its stated purpose of summarizing the best available data concerning executive ability.

WILLIAM O. JENKINS.

*Indiana University.*

BLAIR, GLENN MYERS. *Diagnostic and remedial teaching in secondary schools*. New York: Macmillan, 1946. Pp. xv+422.

Dr. Blair's book has a two-fold purpose: "to supply teachers, principals, supervisors, and superintendents with concrete and practical suggestions for carrying out remedial programs" in secondary schools, and to serve "as a basic text in courses in diagnostic and remedial teaching . . . in teacher-training institutions." Accordingly, the book is divided into three parts. Part I, consisting of the first seven chapters, is devoted to "Diagnostic and Remedial Teaching of Reading." Part 2, consisting of chapters 8 through 11, deals with remedial work in secondary schools, in arithmetic, spelling, handwriting, and the fundamentals of English respectively. Two chapters dealing with the making of case studies and preparation for remedial teaching, comprising Part 3, complete the book.

Dr. Blair is not concerned with the theoretical implications and ramifications of remedial teaching on the secondary level, including even college, as he suggests. He does not treat remedial teaching on the secondary level as a necessary evil, the need for which ought not to exist and which ought to be prevented. Reading, Dr. Blair states "is such a complex skill that it is possible for the elementary school merely to initiate the process and to develop a few of the basic skills" (p. 3). Accordingly, he devotes his first chapter—the introduction—to statements as to the importance of reading, the extent of reading disability on the secondary level, the effects of reading retardation, and the meaning of remedial teaching. Thereafter, he sets forth the procedures, techniques

and plans involved in remedial teaching, in the faith—while admitting the desirability of thorough preparation—that any competent teacher can make a contribution to this area. In the words of Dr. Blair, "For after all, remedial teaching is just good teaching" (pp. 405-406).

The subjects dealt with range from the means of locating the deficient student and the causes of deficiency and their discovery, to the materials and exercises for the improvement of these deficiencies and the manner of employment of such materials and exercises.

A nation-wide survey of remedial teaching in secondary schools, made by the author in the spring of 1940 and discussed in chapter six of the book, formed one of the chief sources for the book's material.

Twenty illustrations and 22 tables, thorough documentation for practically every important fact and/or statement made, and additional extensive references at the end of each chapter, except the last, enhance the value of the volume.

JACOB I. HARTSTEIN.

*Long Island University  
and Yeshiva University.*

DEUTSCH, A. *The mentally ill in America: a history of their care and treatment from colonial times.* New York: Columbia University Press, 1946. Pp. xvii+530.

This book was originally published in 1937 by Doubleday, Doran and Company. After being out of print for some time, it has been re-issued by Columbia University Press. The book is well documented with eighteen pages of bibliography, arranged by chapters. There is also a well-prepared sixteen-page index.

Tracing the history of the care and treatment of mental illness from earliest historical times to the founding of the American colonies, the author presents the colonial period as one of the most shocking periods of our social history as reflected in attitudes toward the mentally ill. From this period, the author traces the historical evolution of changing attitudes and concepts up to the present-day mental hygiene movement, including the story of Clifford Beers, its founder.

The author describes the end of the American Revolution as the real beginnings of American psychiatry under Dr. Benjamin Rush, "the Father of American Psychiatry." Following him, a period of regression ensued until the middle of the nineteenth century when Dorothea Lynde Dix brought the attention of a shocked nation to the care and treatment of the mentally ill and catalyzed it into constructive action. The next outstanding period was the turn of the twentieth century which heralded "the coming of age of psychiatry in America." At this time "asylums" were beginning to be called "hospitals," with the gradual ascendancy of the therapeutic over the custodial ideal, the growth of private psychiatric practice, and the rise of out-patient clinics.

Integrated with this history are chapters on the rise of State responsibility for the mentally ill, glimpses of the outstanding psychiatrists who have contributed to the field, the epic battle between proponents of restraint and non-restraint, the care and treatment of mental deficiency, the criminally insane, laws governing the commitment of the mentally ill, and a final chapter on mental hygiene. This latter chapter provides a dynamic prospectus for the mental hygiene movement and is well worth quoting for its intrinsic value:

A world of peace and freedom, from which the twin specters of war and insecurity will be banished, a world of equal opportunity, where people will be freed from stunting inhibitions and 'guilt feelings' arising from outworn prejudices and taboos, a world where children may lead healthy, happy lives and grow into useful, well-adjusted citizens, where the personality is permitted to develop naturally and freely, where the individual is given a sense of personal worth and dignity, and where his activities and ambitions are integrated with the development of group life—such is the goal toward which mental hygiene must strive.

Psychologists may well take exception to one serious omission in an otherwise well-organized book. Although he recognizes the role that social work has played in this field, practically no recognition is given to psychology's contributions, except for the section on mental defect wherein the author discusses intelligence test development. The author fails to include the part played by psychologists in assisting psychiatrists in diagnosis and treatment through programs of psychological diagnostic testing, vocational and educational counselling, various types of therapy, clinical research and other related activities.

However, the book is an important contribution to the history of mental disease in this country. It charts the path that must be followed if mental hygiene, as well as psychiatry in its true meaning of mental treatment, are to go forward to greater heights of accomplishment rather than to become static or to regress.

JULES D. HOLZBERG

*Connecticut State Hospital*

DUNLAP, KNIGHT. *Personal adjustment*. New York: McGraw-Hill, 1946. Pp. xii+446.

Though apparently meant for the undergraduate and the layman, this book appears to promise little enlightenment on the psychodynamics of maladjusted behavior. The treatment is dogmatic rather than scholarly, as evidenced by the author's choice of expressions, his omission of a bibliography, and his neglect to cite the evidence for atypical generalizations. For example, the reviewer would have appreciated evidence for the statement that "most neurotics are vegetarians and have been vegetarians for the greater part of their lives."



Had the author merely failed to make a contribution to knowledge or understanding, the book would hardly deserve mention. However, the book appears to have had a mission, as stated in the preface, where Dunlap writes that "it is high time that psychoanalysis should be presented in its true light." This purpose does not seem to have been fulfilled, since the author has evidently succeeded only in presenting psychoanalysis from a strongly biased point of view. Few contextual opportunities were missed to malign Freud and psychoanalytic concepts. To the reviewer, it seems that it is no longer fashionable among psychologists to regard psychoanalytic concepts as if they were merely the products of "mythology" and "superstition."

Dunlap's first attempt in the present book to present psychoanalysis "in its true light" states: "The progress that was expected a generation ago to be a resultant of co-operation of psychologists and physicians was thwarted by the rise of psychoanalysis, which is based on ancient popular superstitions." It occurs to the reviewer that the prejudice of American psychologists a generation ago, and their lack at that time of techniques by which to deal effectually with personality dynamics, should share in the responsibility for this failure in co-operation. Probably the public expression by psychologists of attitudes such as those expressed by the author will prolong the difficulties in co-operation which the author deplors.

The reviewer fails to see wherein the total effect of this book can be of service to the psychological profession or to the teaching of psychology. At many points Dunlap states atypical opinions as if they were the generally accepted doctrines of psychologists. The author might have profited by the admonition of modern semanticists that one's own opinions should be so identified rather than being attributed to everybody in possession of the relevant evidence.

In fairness to the author, it should be mentioned that his chapters on sex and marriage are vividly frank in the treatment of a subject which often is veiled or avoided in college texts of this sort. Had the remainder of the book maintained the standard set by these chapters, the reviewer might have foreseen for it a more favorable reception than it deserves in its present form.

BERT R. SAPPENFIELD

*Montana State University*

BAXTER, E. D. *An approach to guidance*. New York: Appleton-Century, 1946. Pp. xii + 305.

This book consists of two parts. Part I is a fictional presentation of the experiences of a director of guidance in a small public school. Her activities, problems, and viewpoints are illustrated in specific incidents and conversations with the school administration, teachers, parents,



and others. Part II is entitled *The Story Interpretation* and attempts to state in concise terms the principles and methods of guidance and education illustrated in the story. There are 138 such interpretations, each clearly related to the story proper by marginal references. Throughout both the story and the story interpretation supporting references are made to recent literature and the volume ends with a 227-item annotated bibliography.

With respect to method of presentation, the book is admirably done. The story is interesting and realistic, the interpretations are to the point, and the bibliography is skillfully annotated—a rare achievement in itself. With respect to content, however, the book can be evaluated fairly only if the orientation and purposes of the author are clearly recognized. In the opinion of the reviewer, the author's attitude toward guidance is best expressed in the following quotations: "every teacher is a counselor" (p. 199) and "adjusting, happy teachers mean adjusting, happy pupils" (p. 3). The "guidance expert" will thus look in vain for a detailed treatment of the use of cumulative records, aptitude tests, home-room programs, administrative charts, etc., such as presented by Traxler, Darley, Williamson, Strang, Reed, and others. Instead he will find emphasis upon inter-personal relationships among pupils, teachers, parents, and administration and upon the place of the guidance director in facilitating these relationships. "Guidance" becomes almost synonymous with "education" and the approach to better education is through the teacher's own adjustment and personality growth and the cultivation of a greater awareness of individual needs.

*An Approach to Guidance* will therefore find its place in the literature on guidance: (1) as a broadening experience to guidance specialists overly engrossed in the mechanical details of administering tests and programs; (2) as a revelation to school administrators of the need for teacher, as well as pupil, guidance; (3) as an inspiration to teachers-in-training; (4) as a general commentary on the goals of education in a democratic society; and (5) as a guide to guidance directors undertaking a new assignment. It is *not* a text book in guidance methods, however, and will serve primarily as supplementary reading in courses on student personnel work, public school guidance and student counseling.

ALBERT S. THOMPSON

*Vanderbilt University*

BLACK, IRMA S. *Off to a good start*. New York: Harcourt, Brace, 1946. Pp. xii+256.

In an informal, often humorous way this "handbook for modern parents" discusses the common problems and perplexities that are an integral part of parenthood. Parents will appreciate not only the sen-

sible and non-technical advice directly given but also its underlying assumption that parents are on the whole a well-meaning lot of people doing a difficult job well. It urges parents to accept and respect their own personalities as well as those of their children, and emphasizes the reassuring fact that there are thousands of ways of being successful either for a parent or a child.

The usual range of fundamental problems is covered and many of the aggravating, if often minor, ones also included. The discussion of the essentials of feeding and habit training is rich in practical suggestions as well as basic in modern theory of the psychological significance of these learnings. Discussion of the child developing as a member of the family group includes good material on discipline, sex instruction, the special problem of the precocious child, as well as the minor worries of birthday parties, Santa Claus and the subtle torments of a week of enforced indoor play. Parts III and IV take the child out into the larger world of relations to other children, to adults from Grandpa to dentist, and the widening horizons of travel and schooling, including an analysis of the progressive school's status. The concluding section on intellectual growth and self-expression gives many helpful suggestions concerning play materials and the problems of readiness for the three R's, together with an evaluation of the roles of radio, movies and circuses in the child's life.

MIRIAM FORSTER FIEDLER

*Vassar College*

## BOOKS AND MATERIALS RECEIVED

ABRAHAMSEN, DAVID. *The mind and death of a genius*. New York: Columbia Univ. Press, 1946. Pp. viii+228.

AKHILANANDER, SWAMI. *Hindu psychology: its meaning for the West*. New York: Harper, 1946. Pp. viii+241.

ARTHUR, GRACE. *Tutoring as therapy*. New York: The Commonwealth Fund, 1946. Pp. ix+125.

BARKER, R. G., WRIGHT, BEATRICE A., AND GONICK, MOLLIE R. *Adjustment to physical handicap: a survey of the social psychology of physique and disability*. Social Science Research Council Bull. 55. New York: Social Science Research Council, 1946. Pp. xi+372.

BAXTER, DOROTHY. *An approach to guidance*. New York: Appleton-Century, 1946. Pp. xii+205.

BRIDGES, J. W. *Psychology—normal and abnormal*. (Rev. Ed.) Toronto: Pitman, 1946. Pp. viii+470.

DELGADO, HONORIO. *La personalidad y el carácter*. Lima, Peru: Talleres Graficos de la Editorial Lumen S. A., 1946. Pp. 207.

FIEANDT, KAI VON. *Yhteisoelämästä Kansanopiston Luontiosilla Opetuskursseilla*. Jyväskylä, Finland: Weilin & Goosin, 1946. Pp. +159.

JELLINCK, E. M. *Phases in the drinking history of alcoholics—analysis of a survey conducted by the Grapevine, official organ of Alcoholics Anonymous*. Memoirs of the Section of Studies on Alcohol, Yale University. New Haven: Hillhouse Press, 1946. Pp. xi+88.

KABACK, GOLDIE RUTH. *Vocational personalities—an application of the Rorschach Group Method*. Teachers College, Columbia Univ., Contr. to Educ., No. 924. New York: Bureau of Publications, Teachers College, Columbia Univ., 1946. Pp. x+116.

KLUCKHOHN, CLYDE, AND LEIGHTON, DOROTHEA. *The Navaho*. Cambridge: Harvard Univ. Press, 1946. Pp. xx+258.

MASE, D. J. *Etiology of articulatory speech defects*. Teachers College, Columbia Univ., Contr. to Educ., No. 921. New York: Bureau of Publications, Teachers College, Columbia Univ., 1946. Pp. viii+85.

MAYS, A. B. *The concept of vocational education in the thinking of the general educator, 1845-1945*. Bureau of Educational Research Bull. No. 62. Urbana, Ill.: College of Education, Univ. of Illinois, 1946. Pp. 107.

PAULSON, BLANCHE. *Wellsprings of action*. Self-Appraisal and Careers Series No. 3. Chicago: Board of Education of the City of Chicago, 1946. Pp. 59.

REIK, THEODORE. *Ritual: psychoanalytic studies*. (The psychological problems of religion, Vol 1.) New York: Farrar, Straus, 1946. Pp. 367.

SMITH, G. MILTON. *A simplified guide to statistics—for psychology and education*. New York: Rinehart, 1946. Pp. xiv+109.

TRAXLER, A. E., AND TOWNSEND, AGATHA. *Another five years of research in reading: summary and bibliography*. Educational Records Bull. No. 46. New York: Educational Records Bureau, 1946. Pp. vi+192.

WITMER, HELEN L. (Ed.) *Psychiatric interviewing with children*. New York: The Commonwealth Fund, 1946. Pp. v+443.

WOLFF, WERNER. *The personality of the pre-school child*. New York: Grune & Stratton, 1946. Pp. xvi+341.

*Forty-first annual report*. The Carnegie Foundation for the Advancement of Teaching (1945-1946.) New York: The Carnegie Foundation for the Advancement of Teaching, 1946. Pp. 160.

*How to make paper-and-pencil tests*. Personnel Research Section, The Adjutant General's Office, Classification and Replacement Branch. New York: The Adjutant General's Office, 1945. Pp. 44.

*Personnel classification tests*. War Department Technical Manual 12-260. Washington: Government Printing Office, 1946. Pp. iv+90.

*Selection and employment of transit operating personnel*. A set of five manuals prepared by Glenn U. Cleeton *et al.* New York: American Transit Ass., 1946.







